F1000Research

Check for updates

BRIEF REPORT

Use of the informational spectrum methodology for rapid biological analysis of the novel coronavirus 2019-nCoV: prediction of potential receptor, natural reservoir, tropism and therapeutic/vaccine target [version 1; peer review: awaiting peer review]

Veljko Veljkovic [ID] [1], Júlia Vergara-Alert[2], Joaquim Segalés[3,4], Slobodan Paessler[5]

[1]Biomed Protection, Galveston, TX, 77550, USA
[2]IRTA, Centre de Recerca en Sanitat Animal (CReSA, IRTA-UAB), Campus de la Universitat Autònoma de Barcelona, Barcelona, 08193 Bellaterra, Spain
[3]Departament de Sanitat i Anatomia Animals, Universitat Autònoma de Barcelona (UAB), Barcelona, 08193 Bellaterra, Spain
[4]UAB, Centre de Recerca en Sanitat Animal (CReSA, IRTA-UAB), Campus de la Universitat Autònoma de Barcelona, Barcelona, 08193 Bellaterra, Spain
[5]Department of Pathology, Galveston National Laboratory, University of Texas Medical Branch, Galveston, TX, 77555, USA

Abstract
A novel coronavirus recently identified in Wuhan, China (2019-nCoV) has expanded the number of highly pathogenic coronaviruses affecting humans. The 2019-nCoV represents a potential epidemic or pandemic threat, which requires a quick response for preparedness against this infection. The present report uses the informational spectrum methodology to identify the possible origin and natural host of the new virus, as well as putative therapeutic and vaccine targets. The performed in silico analysis indicates that the newly emerging 2019-nCoV is closely related to severe acute respiratory syndrome (SARS)-CoV and, to a lesser degree, Middle East respiratory syndrome (MERS)-CoV. Moreover, the well-known SARS-CoV receptor (ACE2) might be a putative receptor for the novel virus as well. Additional results indicated that civets and poultry are potential candidates for the natural reservoir of the 2019-nCoV, and that domain 288-330 of S1 protein from the 2019-nCoV represents promising therapeutic and/or vaccine target.

Keywords
2019-nCoV, Wuhan coronavirus, SARS, MERS
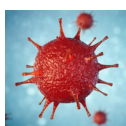
This article is included in the Disease Outbreaks gateway.

Open Peer Review

Reviewer Status ✓

	Invited Reviewers
	1
version 2 (update) 31 Jan 2020	✓ report
version 1 27 Jan 2020	

1 **Rita Casadio**, University of Bologna, Bologna, Italy

Any reports and responses or comments on the article can be found at the end of the article.

Corresponding author: Veljko Veljkovic (eveljkve@gmail.com)

Author roles: Veljkovic V: Conceptualization, Formal Analysis, Methodology, Writing – Original Draft Preparation; **Vergara-Alert J**: Investigation, Writing – Original Draft Preparation; **Segalés J**: Investigation, Supervision, Writing – Original Draft Preparation; **Paessler S**: Investigation, Supervision, Writing – Original Draft Preparation

Introduction

Fears are mounting worldwide over the cross-border spread of the new strain of coronavirus (denoted as 2019-nCoV) originated in Wuhan, the largest city in central China, after its spread to Thailand and Japan. The newly emerging pathogen belongs to the same virus family as the deadly severe acute respiratory syndrome and Middle East respiratory syndrome coronaviruses (SARS-CoV and MERS-CoV, respectively). The World Health Organization (WHO) has recently published surveillance recommendations for a possible "large epidemic or even pandemic" of the novel coronavirus and it has issued guidelines for hospitals across the world. However, many questions about 2019-nCov remain unanswered: (i) what is the origin and/or natural reservoir of the virus? (ii) is it easily transmitted from human to human? and (iii) what are the potential diagnostic, therapeutic and vaccine targets? Currently, only nucleotide sequences of eight human 2019-nCoV isolates are available without any additional information about biological properties of the virus, beyond the morphology confirmation of the virion using electronic microscopy. This is likely not enough information to answer the important abovementioned questions.

The informational spectrum method (ISM), a virtual spectroscopy method for analysis of proteins, is based on the fundamental electronic properties of amino acids and requires only nucleotide sequence availability to investigate proteins[1]. For this reason, ISM was previously used for analysis of novel viruses for which little or no information were available[2–5]. Here, the 2019-nCoV was analyzed with ISM to identify its possible origin and natural host, as well as putative therapeutic and vaccine targets.

Methods

Sequences

The S1 surface protein sequences from 8 human 2019-nCoV, deposited in the publicly available GISAID database (assessed on January 19, 2020), were analyzed by ISM. The studied sequences were BetaCoV/Wuhan/IVDC-HB-04/2020, BetaCoV/Wuhan/IVDC-HB-01/2019, BetaCoV/Wuhan/IVDC-HB-05/2019, BetaCoV/Wuhan/IPBCAMS-WH-01/2019, BetaCoV/Wuhan/WIV04/2019, BetaCoV/Wuhan-Hu-1/2019, BetaCoV/Nonthaburi/61/2020, and BetaCoV/Nonthaburi/74/2020.

In the phylogenetic analysis, different amino acid sequences of other coronaviruses were also included: (i) S1 proteins from the following viruses: AVP78042, AVPvp78031, AY304486, AY559093, JX163927, YN2018B, KY417146, used already by other authors in the study of the phylogenetic relationship between 2019-nCoV and nearest bat and SARS-like CoVs (GISAID database); and (ii) S1 proteins from three first isolated human MERS-CoV: AGG22542, AFS88936, AFY13307, deposited in the GISAID database

The ISM

Detailed description of the sequence analysis based on ISM has been published elsewhere[2]. According to this approach, sequences (protein or DNA) are transformed into signals by assignment of numerical values of each element (amino acid or nucleotide). These values correspond to electron-ion interaction potential[6], determining electronic properties of amino acid/nucleotides, which are essential for their intermolecular interactions. The signal obtained is then decomposed in a periodical function by the Fourier transformation. The result is a series of frequencies and their amplitudes. The obtained frequencies correspond to the distribution of structural motifs (primary structure) with defined physico-chemical characteristics responsible for the biological function of the putative protein corresponding to the analyzed sequence. When comparing proteins that share same biological or biochemical function, the technique allows detection of code/frequency pairs that are specific for their common biological properties. The method is insensitive to the location of the motifs and, therefore, does not require previous alignment of the sequences. In addition, this is the only method that allows immediate functional analysis.

Phylogenetic analysis

The phylogenetic tree of S1 proteins from coronaviruses was generated with the ISM-based phylogenetic algorithm ISTREE, previously described in detail elsewhere[7]. In the presented analysis, we calculated the distance matrix with the amplitude on the frequency F(0.257) as the distance measure between sequences.

Results and discussion

In order to compare informational similarity between 2019-nCoV, SARS-CoV, MERS-CoV and Bat SARS-like CoV, the cross-spectra (CS) of S1 proteins from these viruses were calculated. Figure 1a shows the CS of 2019-nCoV, SARS-CoV and MERS-CoV. These CS contain only one dominant peak corresponding to the frequency F(0.257). Figure 1b displays the CS of S1 proteins from 2019-nCoV and Bat SARS-like CoV. Amplitudes in these latter CS are significantly lower than in those CS presented in Figure 1a. These results show that (i) S1 proteins from 2019-nCoV, SARS-CoV, MERS-CoV and Bat SARS-like CoV encode common information, which is represented with the frequency F(0.257), and (ii) S1 proteins from 2019-nCoV are remarkable more informationally similar with S1 from SARS-CoV and MERS-CoV than with S1 from Bat SARS-like CoV. This suggests that biological properties of 2019-nCoV are apparently more similar to SARS-CoV and MERS-CoV than to Bat SARS-like CoV.

To confirm this conclusion, the ISM-base phylogenetic tree for S1 proteins was calculated (Figure 2). In this calculation the amplitude on the frequency F(0.257) was used as the distance measure. As observed in Figure 2, all analyzed 2019-nCoV S1 amino acid sequences are grouped with SARS-CoV and MERS-CoV and separated from Bat SARS-like CoV. This indicates that 2019-nCoV are more phylogenetically similar to SARS-CoV and MERS-CoV than to Bat SARS-like CoV. This result differs from those obtained with the homology-based phylogenetic analysis, which showed that 2019-CoV are closely related to Bat SARS-like CoV (https://platform.gisaid.org/epi3/frontend#lightbox1296857287).
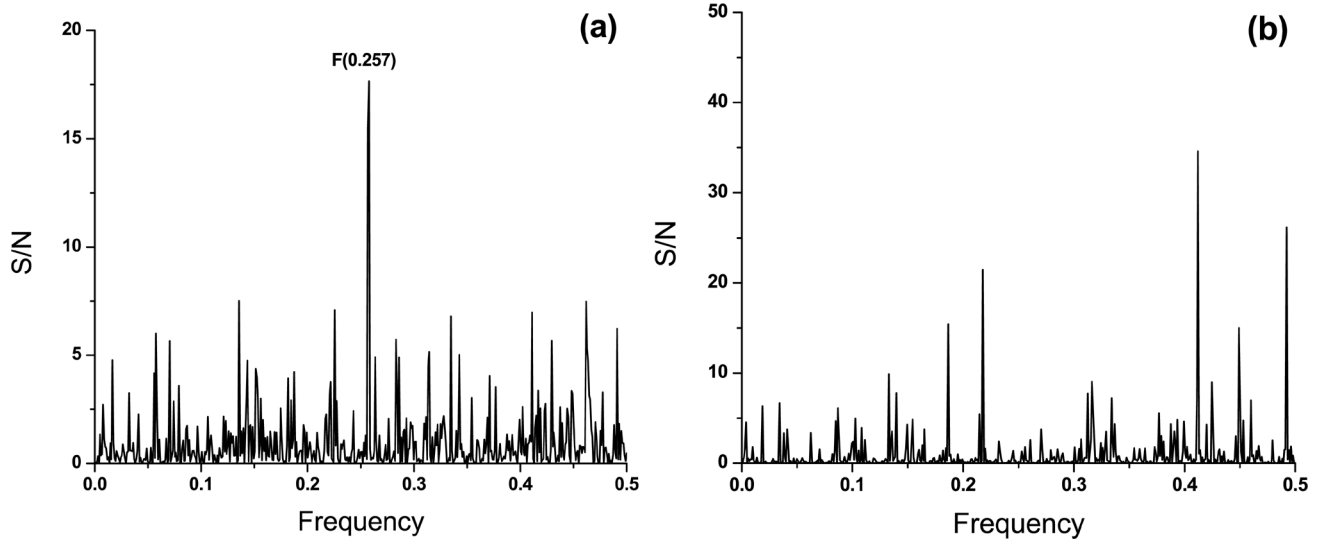
Figure 1. Cross-spectrum (CS) of S1 proteins. (**a**) CS of S1 from SARS-CoV, MERS-CoV and 2019-nCoV; (**b**) CS of Bat SARS-like CoV and 2019-NCov. The abscissa represents the frequencies from the Fourier transform of the sequence of electron-ion interaction potential corresponding to the amino-acid sequence of proteins. The lowest frequency is 0.0 and the highest is 0.5. The ordinate represents the signal-to-noise ratio (the ratio between signal intensity at one particular IS frequency and the main value of the whole spectrum, S/N).
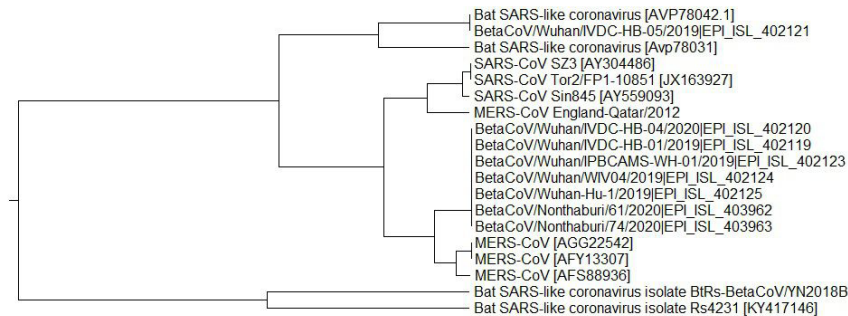

Figure 2. Informational spectrum method-based phylogenetic tree for S1 proteins from SARS-CoV, MERS-CoV, Bat SARS-like CoV and 2019-nCoV. The frequency F(0.257) as the distance measure was used.

It has been previously shown that the dominant frequency in the informational spectrum of viral envelope proteins corresponds to interaction between the virus and its receptor[2,3,7,8]. The ISM analysis showed that the frequency component F(0.257) is present in the CS of S1 SARS-CoV and its receptor angiotensin converting enzyme 2 (ACE2)[9], but not in the CS of S1 MERS-CoV and its main receptor dipeptidyl peptidase 4 (DPP4)[10]. Of note is that both receptors ACE2 and DPP4 are expressed in airway epithelia. Presence of F(0.257) in the informational spectrum of MERS-CoV (Figure 1) suggests also possible interaction between this virus and the ACE2. The dominant peak on the frequency F(0.257) in the CS of S1 from SARS-CoV and MERS-CoV and ACE2 supports this possibility (Figure 3), although this has not been formally proved for MERS-CoV[11].

As it is shown in Figure 1a, the frequency F(0.257) is also present in the informational spectrum of the 2019-nCoV, suggesting that ACE2 might be the receptor for this novel coronavirus too. Calculation of the CS for S1 protein from the 2019-nCoV and all ACE2 sequences available at the UniProt database revealed that the highest amplitudes on the frequency F(0.257) correspond to ACE2 from civet and chicken. This result indicates that these species can be included as potential candidates for the natural reservoir of the 2019-nCoV. However, it is possible that 2019-nCoV viruses use very different receptors in the natural host(s) and not only the ACE2 as it is the putative case in humans.

Finally, the S1 amino acid sequence from the 2019-nCoV was scanned to look for the domain that gives the highest contribution to the information represented by the frequency F(0.257) (Figure 4a). This analysis revealed domain 266–330 (numbering concerns the maturated protein) is essential for interaction of 2019-nCoV with ACE2. Of note is the striking homology between these domains of S1 proteins from 2019-nCoV and SARS-CoV, but not from MERS-CoV for which ACE2 is not the main receptor (Figure 4b).
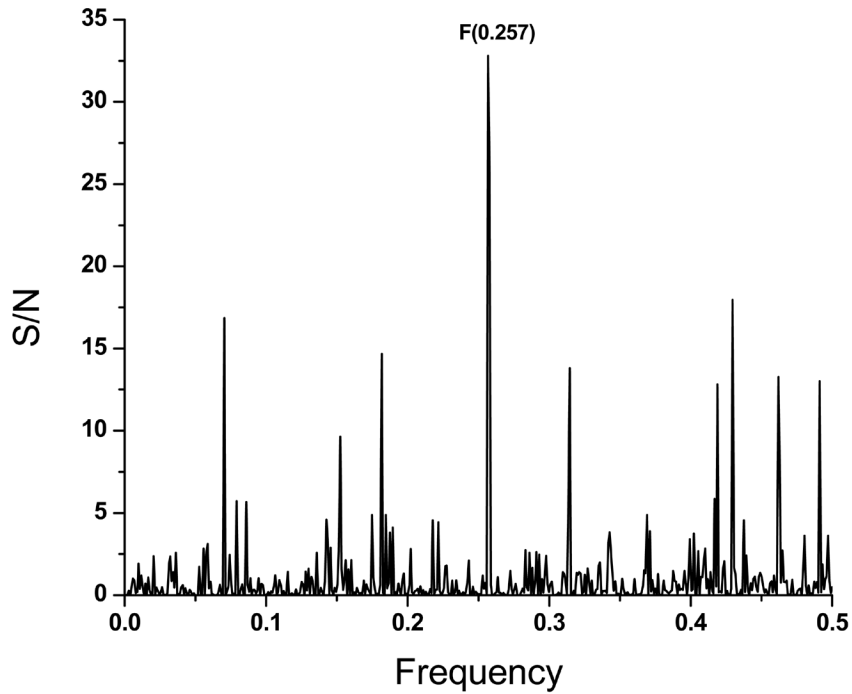
Figure 3. Cross-spectrum of ACE2 and S1 proteins from SARS-CoV and MERS-CoV. The abscissa and the ordinate are as described in Figure 1.
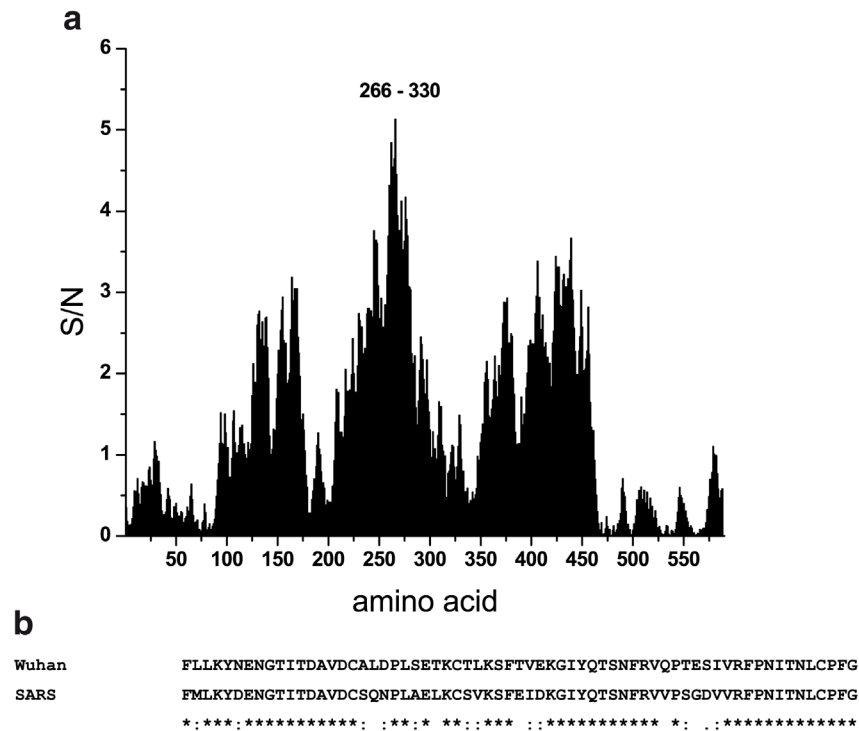
Figure 4. Domain of S1 protein which is important for 2019-nCoV/ACE2 interaction. (**a**) Mapping of the domain of S1 protein from 2019-nCoV (BetaCoV/Wuhan/IVDC-HB-01/2019) which gives the dominant contribution to the information represented with the frequency F(0.257). (**b**) Sequence homology between domains of S1 proteins from SARS-CoV and 2019-nCoV with essential contribution to the information corresponding to the frequency F(0.257).

In conclusion, results of the presented *in silico* analysis suggest the following: (i) the newly emerging 2019-nCoV is highly related to SARS-CoV and, to a lesser degree, MERS-CoV; (ii) civets and poultry are potential candidates for the natural reservoir of the 2019-nCoV and (iii) domain 288–330 of S1 protein from the 2019-nCoV represents promising therapeutic and/or vaccine target. Further research on these issues are needed, including the development of reverse genetics and animal models to study the biology of 2019-nCoV.

Data availability
Underlying data
Sequence data of the viruses were obtained from the GISAID EpiFlu™ Database. To access the database each individual user should complete the "Registration Form For Individual Users", which is available alongside detailed instructions. After submission of the Registration form, the user will receive a password. There are not any other restrictions for the access to GISAID. Conditions of access to, and use of, the GISAID EpiFlu™ Database and Data are defined by the Terms of Use.

References

1. Veljković V, Cosić I, Dimitrijević B, *et al.*: **Is it possible to analyze DNA and protein sequences by the methods of digital signal processing?** *IEEE Trans Biomed Eng.* 1985; **32**(5): 337–41.
   PubMed Abstract | **Publisher Full Text**

2. Veljkovic V, Niman HL, Glisic S, *et al.*: **Identification of hemagglutinin structural domain and polymorphisms which may modulate swine H1N1 interactions with human receptor.** *BMC Struct Biol.* 2009; **9**: 62 .
   PubMed Abstract | **Publisher Full Text** | **Free Full Text**

3. Veljkovic V, Glisic S, Muller CP, *et al.*: *In silico* **analysis suggests interaction between Ebola virus and the extracellular matrix.** *Front Microbiol.* 2015; **6**: 135.
   PubMed Abstract | **Publisher Full Text** | **Free Full Text**

4. Paessler S, Veljkovic V: **Prediction of influenza vaccine effectiveness for the influenza season 2017/18 in the US [version 1; peer review: 2 approved].** *F1000Res.* 2017; **6**: 2067.
   PubMed Abstract | **Publisher Full Text** | **Free Full Text**

5. Koma T, Veljkovic V, Anderson DE, *et al.*: **Zika virus infection elicits auto-antibodies to C1q.** *Sci Rep.* 2018; **8**(1): 1882.
   PubMed Abstract | **Publisher Full Text** | **Free Full Text**

6. Veljković V, Slavić I: **Simple general-model pseudopotential.** *Phys Rev Let.* 1972;

   29: 105.
   Publisher Full Text

7. Perovic VR, Muller CP, Niman HL, *et al.*: **Novel phylogenetic algorithm to monitor human tropism in Egyptian H5N1-HPAIV reveals evolution toward efficient human-to-human transmission.** *PLoS One.* 2013; **8**(4): e61572.
   PubMed Abstract | **Publisher Full Text** | **Free Full Text**

8. Schmier S, Mostafa A, Haarmann T, *et al.*: *In Silico* **Prediction and Experimental Confirmation of HA Residues Conferring Enhanced Human Receptor Specificity of H5N1 Influenza A Viruses.** *Sci Rep.* 2015; **5**: 11434.
   PubMed Abstract | **Publisher Full Text** | **Free Full Text**

9. Li W, Moore MJ, Vasilieva N, *et al.*: **Angiotensin-converting enzyme 2 is a functional receptor for the SARS coronavirus.** *Nature.* 2003; **426**(6965): 450–4.
   PubMed Abstract | **Publisher Full Text**

10. Raj VS, Mou H, Smits SL, *et al.*: **Dipeptidyl peptidase 4 is a functional receptor for the emerging human coronavirus-EMC.** *Nature.* 2013; **495**(7440): 251–4.
    PubMed Abstract | **Publisher Full Text**

11. Müller MA, Raj VS, Muth D, *et al.*: **Human coronavirus EMC does not require the SARS-coronavirus receptor and maintains broad replicative capability in mammalian cell lines.** *mBio.* 2012; **3**(6): pii: e00515-12.
    PubMed Abstract | **Publisher Full Text** | **Free Full Text**

The benefits of publishing with F1000Research:

- Your article is published within days, with no editorial bias

- You can publish traditional articles, null/negative results, case reports, data notes and more

- The peer review process is transparent and collaborative

- Your article is indexed in PubMed after passing peer review

- Dedicated customer support at every stage

For pre-submission enquiries, contact research@f1000.com

F1000Research