



Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.

Journal Pre-proof

The Role of Phylogenetic Analysis in Clarifying the Infection Source of a COVID-19 Patient

Jann-Tay Wang , You-Yu Lin , Sui-Yuan Chang , Shiou-Hwei Yeh , Bor-Hsian Hu , Pei-Jer Chen , Shan-Chwen Chang

PII: S0163-4453(20)30159-6
DOI: <https://doi.org/10.1016/j.jinf.2020.03.031>
Reference: YJINF 4505



To appear in: *Journal of Infection*

Accepted date: 21 March 2020

Please cite this article as: Jann-Tay Wang , You-Yu Lin , Sui-Yuan Chang , Shiou-Hwei Yeh , Bor-Hsian Hu , Pei-Jer Chen , Shan-Chwen Chang , The Role of Phylogenetic Analysis in Clarifying the Infection Source of a COVID-19 Patient, *Journal of Infection* (2020), doi: <https://doi.org/10.1016/j.jinf.2020.03.031>

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2020 The British Infection Association. Published by Elsevier Ltd. All rights reserved.

The Role of Phylogenetic Analysis in Clarifying the Infection Source of a COVID-19 Patient

Jann-Tay Wang^{1*}, You-Yu Lin^{2*}, Sui-Yuan Chang^{3, 4*}, Shiou-Hwei Yeh^{2*}, Bor-Hsian
Hu⁵, Pei-Jer Chen², Shan-Chwen Chang^{1, 6}

¹Department of Internal Medicine, National Taiwan University Hospital, Taipei,
Taiwan

²Department of Microbiology, National Taiwan University Center of Genomic
Medicine, Taipei, Taiwan

³Department of Clinical Laboratory Sciences and Medical Biotechnology, College of
Medicine, National Taiwan University, Taipei, Taiwan

⁴Department of Laboratory Medicine, National Taiwan University Hospital, Taipei,
Taiwan

⁵Department of Internal Medicine, Hoping Fuyou Branch of Taipei City Hospital,
Taipei, Taiwan

⁶Department of Internal Medicine, College of Medicine, National Taiwan University,
Taipei, Taiwan

*Contribute equally

Corresponding author information

Shan-Chwen Chang, M.D., Ph.D.

Department of Internal Medicine, National Taiwan University Hospital, No.7,
Chung-Shan South Rd., Taipei City 10002, Taiwan (R.O.C.).

E-mail address: changsc@ntu.edu.tw

Tel: +886-2-23123456 ext. 88799

Fax:+886-2-23971412.

Journal Pre-proof

Dear Editor,

Previous reports indicated that the emergence of the novel coronavirus (SARS-CoV-2) infection (COVID-19) had raised global concern and was characterized as a pandemic event by the World Health Organization on March 11, 2020 [1-3]. Till March 18, 2020, it has spread to 146 countries, including Taiwan [4]. People in Taiwan and mainland China travel frequently, which put Taiwan at a great risk of acquiring an epidemic of COVID-19. Taiwan has been on constant alert and react rapidly to epidemics change from China ever since the severe acute respiratory syndrome (SARS) epidemic in 2003 and has done much effort on the containment of COVID-19 with success [5]. Till March 18, 2020, there are only 100 cases of COVID-19 noted in Taiwan, including 79 imported cases and 21 cases belonging to seven occasions of limited local transmission (six family clusters and two transmissions in social societies) [4].

To contain the epidemics of COVID-19, prevention from both import and export of contagious people is an essential intervention. It is also important to clearly clarify the infection source in order to initiate an efficient and successful contact tracing for the SARS-CoV-2 infected patients, and thereafter people exposed to the contagious patients could be quarantined to avoid further disease spreading. Here we present a COVID-19 patient whose infection source could not be completely clarified

initially, and later was illuminated by using the phylogenetic analysis of the isolated virus.

This 66-year-old Taiwanese woman was well before. She traveled to Dubai from January 29 to February 10, 2020, and Egypt from February 11 to February 21, 2020. When she stayed in Egypt, she ever participated an eight-day tourism on a Nile cruise boat. She returned to Taiwan via an international airline on February 21, 2020. She began to suffer from general malaise, myalgia, cold sweating, productive cough, and sore throat since February 18, 2020. She reported that there were another 16 persons in the same tourism group had similar symptoms at that time. Her symptoms persisted despite medication, prescribed on February 21 by a local medical doctor. On February 26, the cough exacerbated, and she developed chest tightness, abdominal upset and vomiting. She visited the Department of Emergency of a teaching hospital in Taipei on February 28. No fever was noted during her disease course. In the context of her travel history and prominent respiratory symptoms, a nasopharyngeal swab was taken for test of SARS-CoV-2 by real-time reverse transcription-polymerase chain reaction (RT-PCR), and the result was positive. She was then transferred to a negative-pressure isolation room as a case of COVID-19.

The most interesting point of this patient is where she contracted her COVID-19. By history, she is more likely to contract SARS-CoV-2 infection while travelling

abroad. However, despite that the median incubation period of COVID-19 was 5.1 days, it might be as long as more than three weeks in some extreme cases [6, 7]. Therefore, an argument that she got the infection while she was in Taiwan couldn't be excluded completely.

To clarify this argument, more virologic studies were conducted. Virus whole genome sequencing was conducted for the SARS-CoV-2 isolate (NTU03) from her throat swab collected on March 2, 2020. The derived NTU03 sequence was most similar to clade A2a with only 5 nucleotide differences, which included 2 synonymous mutations (Orf 1b/5410 CTA>TTA, and orf 3a/819 GTG>GTT), 2 nonsynonymous mutations (Orf 1b/799 G>V, and orf 3a/57Q>H), and a mutation within 5'UTR (Table). An average of 12 nucleotide differences were observed between NTU03 and sequences of other clades, whereas an average of 10 nucleotide differences were observed with other previous viruses isolated from Taiwan. The phylogenetic analysis also reveals that the NTU03 belongs to clade A2a (Figure).

Based on the results of whole genome sequencing and phylogenetic analysis, NTU03 belongs to clade A2a, in which all other of the reported case patients were currently either from Europe or travelled to Europe recently according to the information provided by the laboratories who submitted the clade A2a sequences to the Global Initiative on Sharing All Influenza Data (GISAID). None of the previously

submitted sequences of viruses isolated from Taiwan were assigned to clade A2a or A2.

With the limited transmission clusters in Taiwan and the fact that NTU03 exhibits at least 8 unique nucleotide difference compared to other previously reported viruses from Taiwan, we conclude that it is much more probable that the present patient was infected during her travelling abroad. To our best knowledge, in late February and early March, several foreign COVID-19 patients who had travel histories to Egypt were reported. Further phylogenetic analysis including viral sequences derived from the 45 confirmed SARS-CoV-2 infections on the quarantined Nile cruise boat will help to delineate the outbreak on Nile tourism boat and its impact on the COVID-19 epidemic [8].

Acknowledgments: We thank all the persons involved in the response to this outbreak.

Conflict of interest: none to declare.

Financial support: This study was financially supported by grant from the “Center of Precision Medicine” from The Featured Areas Research Center Program within the framework of the Higher Education Sprout Project by the Ministry of Education (MOE) in Taiwan.

References

1. Tang JW, Tambyah PA, Hui DSC. Emergence of a novel coronavirus causing respiratory illness from Wuhan, China. *J Infect.* 2020;80(3):350-371.
2. Zhang Z, Xiao K, Zhang X, Roy A, Shen Y. Emergence of SARS-like Coronavirus in China: An Update. *J Infect.* 2020 doi: 10.1016/j.jinf.2020.03.010. [Epub ahead of print].
3. World Health Organization [Internet]. Statement on the second meeting of the International Health Regulations (2005) Emergency Committee regarding the outbreak of novel coronavirus (2019-nCoV). [cited 2020 Mar 18]. Available from: <https://www.who.int/emergencies/diseases/novel-coronavirus-2019/events-as-they-happen>.
4. Taiwan Centers for Disease Control. Situation of COVID-19 in Taiwan. [cited 2020 Mar 18]. Available from: <https://sites.google.com/cdc.gov.tw/2019ncov/taiwan>.
5. Wang CJ, Ng CY, Brook RH. Response to COVID-19 in Taiwan: Big data analytics, new technology, and proactive testing. *JAMA.* 2020 Mar 3. doi: 10.1001/jama.2020.3151. [Epub ahead of print].
6. Backer JA, Klinkenberg D, Wallinga J. Incubation period of 2019 novel

- coronavirus (2019-nCoV) infections among travelers from Wuhan, China, 20028
January 2020. *Euro Surveill.* 2020;25(5):2000062.
7. Lauer SA, Grantz KH, Bi Q, Jones FK, Zheng Q, Meredith HR, et al. The Incubation Period of Coronavirus Disease 2019 (COVID-19) From Publicly Reported Confirmed Cases: Estimation and Application. *Ann Intern Med.* 2020 Mar 10. doi: 10.7326/M20-0504. [Epub ahead of print].
 8. The Washington Post. Egypt tries to reassure tourists as coronavirus spikes. [cited 2020 Mar 18]. Available from: https://www.washingtonpost.com/world/middle-east/egypt-tries-to-reassure-tourists-as-coronavirus-spikes/2020/03/08/ab1dfbf6-6160-11ea-8a8e-5c5336b32760_story.html.
 9. Hadfield J, Megill C, Bell SM, Huddleston J, Potter B, Callender C, et al. Nextstrain: real-time tracking of pathogen evolution. *Bioinformatics.* 2018;34(23):4121-23.
 10. Sagulenko P, Puller V, Neher RA. TreeTime: Maximum-likelihood phylodynamic analysis. *Virus Evol.* 2018;4(1):vex042.

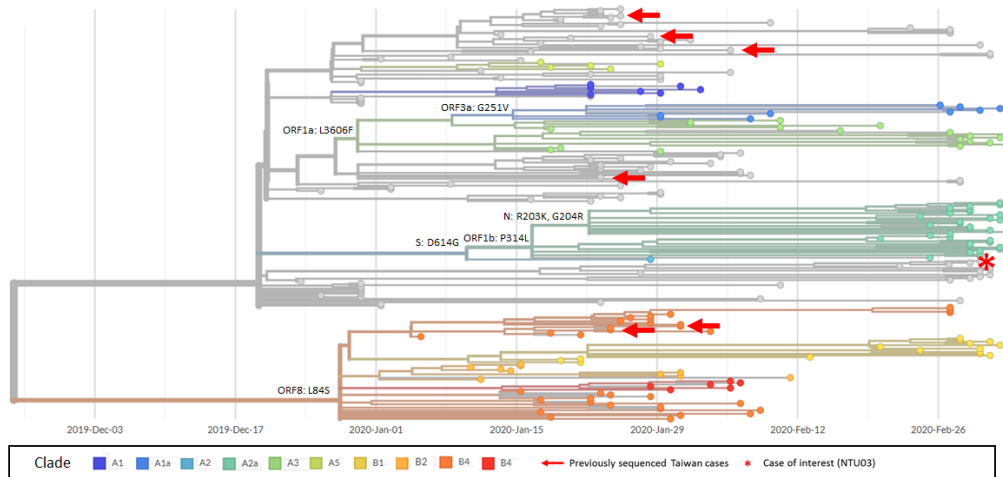


Figure. Phylogenetic analysis of the full-length SARS-CoV-2 sequences. The phylogeny tree analysis was conducted to determine the clade of NTU03 (red asterisk) and its relationship to other viral sequences derived from case patients identified in Taiwan (red arrows). The phylogenetic tree was generated and modified for display purposes from Nextstrain (<https://nextstrain.org/ncov>) [9], which uses genetic sequences and metadata from GISAID (<https://www.gisaid.org/CoV2020/>) and sequence submission date for the horizontal axis [10]. The phylogenetic tree was generated at 2020/03/09 6PM (GMT+8) with a total of 240 viral genomes sampled.

Table. Summary of SARS-CoV-2 alignment

MN908947 position	MN	A1a	A1	A2a	A2	A3	A5	B1	B2	B4-1	B4-2	NTU01	NTU02	NTU03	CDC02	CDC03	CDC04	CGMH1	CDS	CDS position	Codon	Amino acid	
187	A	A	A	G	A	n/a	A	A	A	A	A	A	A	A	A	A	A	A	A	5'UTR	n/a	n/a	n/a
241	C	C	C	T	T	n/a	C	C	C	C	C	C	C	T	C	C	C	C	C	5'UTR	n/a	n/a	n/a
1397	G	G	G	G	G	A	G	G	G	G	G	G	G	G	G	G	G	G	G	orf1a	1132	GTA>ATA	378 V>I
2091	C	C	C	C	C	C	T	C	C	C	C	C	C	C	C	C	C	C	C	orf1a	1826	ACT>ATT	609 T>I
2113	C	C	C	C	C	T	C	C	C	C	C	C	C	C	C	C	C	C	C	orf1a	1848	ATC>ATT	616 I>I
3037	C	C	C	T	T	C	C	C	C	C	C	C	C	T	C	C	C	C	C	orf1a	2772	TTC>TTT	924 F>F
4402	T	T	T	T	T	T	T	T	T	C	T	T	T	T	T	T	T	T	T	orf1a	4137	CTT>CTC	1379 L>L
5062	G	G	G	G	G	G	G	G	G	T	G	G	G	G	G	G	G	G	G	orf1a	4797	TTG>TTT	1599 L>F
8782	C	C	C	C	C	C	C	T	T	T	T	T	C	C	C	T	C	C	C	orf1a	8517	AGC>AGT	2839 S>S
9034	A	A	A	A	A	A	A	A	A	A	A	A	G	A	A	A	A	A	A	orf1a	8769	AAA>AAG	2923 K>K
9430	C	C	C	C	C	C	C	A	C	C	C	C	C	C	C	C	C	C	C	orf1a	9165	ATC>ATA	3055 I>I
9491	C	C	C	C	C	C	C	C	C	C	C	C	T	C	C	C	C	C	C	orf1a	9226	CAT>TAT	3076 H>Y
11083	G	T	G	G	G	T	G	G	G	G	G	G	G	G	G	G	G	G	G	orf1a	10818	TTG>TTT	3606 L>F
13679	A	A	G	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	orf1b	212	TAC>TGC	71 Y>C
14408	C	C	C	T	C	C	C	C	C	C	C	C	C	T	C	C	C	C	C	orf1b	941	CCT>CTT	314 P>L
14805	C	T	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	orf1b	1338	TAC>TAT	446 Y>Y
15863	C	C	C	C	C	C	C	C	C	C	C	C	C	Y	C	C	C	C	C	orf1b	2396	GGA>GTA	799 G>V
16188	G	G	G	G	G	G	G	G	G	G	G	G	G	G	T	G	G	G	G	orf1b	2721	TGG>TGT	907 W>C

17247	T	C	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	orf1b	3780	CGT>CGC	1260 R>R
17373	C	C	T	C	C	C	C	C	C	C	C	C	C	C	C	C	C	orf1b	3906	GCC>GCT	1302 A>A
17747	C	C	C	C	C	C	C	T	C	C	C	C	C	C	C	C	C	orf1b	4280	CCT>CTT	1427 P>L
17858	A	A	A	A	A	A	A	G	A	A	A	A	A	A	A	A	A	orf1b	4391	TAT>TGT	1464 Y>C
18060	C	C	C	C	C	C	C	T	C	C	C	C	C	C	C	C	C	orf1b	4593	CTC>CTT	1531 L>L
18603	T	T	T	T	T	T	T	T	C	T	T	T	T	T	T	T	T	orf1b	5136	CAT>CAC	1712 H>H
18877	C	C	C	C	C	C	C	C	C	C	C	C	C	T	C	C	C	orf1b	5410	CTA>TTA	1804 L>L
18928	C	C	C	C	C	T	C	C	C	C	C	C	C	C	C	C	C	orf1b	5461	CCT>TCT	1821 P>S
18975	T	T	T	T	T	T	T	T	A	T	T	T	T	T	T	T	T	orf1b	5508	GTT>GTA	1836 V>V
19175	A	A	A	A	A	A	A	A	C	A	A	A	A	A	A	A	A	orf1b	5708	GAT>GCT	1903 D>A
21707	C	C	C	C	C	C	T	C	C	C	C	C	C	C	C	C	C	S	145	CAT>TAT	49 H>Y
23403	A	A	A	G	G	A	A	A	A	A	A	A	A	G	A	A	A	S	1841	GAT>GGT	614 D>G
24378	C	T	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	S	2816	TCT>TTT	939 S>F
25563	G	G	G	G	G	G	G	G	G	G	G	G	G	T	G	G	G	orf3a	171	CAG>CAT	57 Q>H
25964	A	A	A	A	A	A	A	A	A	A	A	A	A	A	G	A	A	orf3a	572	GAA>GGA	191 E>G
26144	G	T	G	G	G	G	G	G	G	G	G	G	G	G	T	G	G	orf3a	752	GGT>GTT	251 G>V
26211	G	G	G	G	G	G	G	G	G	G	G	G	G	K	G	G	G	orf3a	819	GTG>GTT	273 V>V
26894	C	T	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	M	372	CTC>CTT	124 L>L
27925	C	C	C	C	C	C	C	T	C	C	C	C	C	C	C	C	C	orf8	32	ACA>ATA	11 T>I
28144	T	T	T	T	T	T	T	C	C	C	C	C	T	T	T	C	T	orf8	251	TTA>TCA	84 L>S
28688	T	T	T	T	T	C	T	T	T	T	T	T	T	T	T	T	T	N	415	TTG>CTG	139 L>L
28878	G	G	G	G	G	G	G	G	G	A	G	G	G	G	G	G	G	N	605	AGT>AAT	202 S>N

29095	C	C	C	C	C	C	C	C	T	C	C	C	C	C	C	C	C	C	N	822	TTC>TTT	274 F>F
29374	G	G	G	G	G	A	G	G	G	G	G	G	G	G	G	G	G	G	N	1101	GAG>GAA	367 E>E
29742	G	G	G	G	G	n/a	G	G	G	G	A	n/a	G	G	G	G	G	G	3'UTR	n/a	n/a	n/a

SARS-CoV-2 genetic sequences, including NTU03, all 6 previously submitted Taiwan sequences, and representative sequences for clades A1, A1a, A2, A2a, A3, A5, B1, B2, and B4, were aligned and compared. MN908947 was used as the reference sequence, and orange shaded nucleotides indicated nucleotides different to the reference nucleotide, and grey shaded positions indicated lack of nucleotide information. The NTU01 and CDC03 are derived from the Case 3 patient, and CDC04 and CGMH1 are from the from the Case 4 patient. Sequence data were obtained from GISAID (<https://www.gisaid.org/CoV2020/>). The GISAID accession number for each representative clade sequence: A1a, hCoV-19/Switzerland/1000477102/2020|EPI_ISL_413019; A1, hCoV-19/Singapore/11/2020|EPI_ISL_410719; A2a, hCoV-19/Italy/UniSR1/2020|EPI_ISL_413489; A2, hCoV-19/Germany/BavPat1/2020|EPI_ISL_406862; A3, hCoV-19/Australia/NSW13/2020|EPI_ISL_413599; A5, hCoV-19/USA/CA5/2020|EPI_ISL_408010; B1, hCoV-19/USA/WA12-UW8/2020|EPI_ISL_413563; B2, hCoV-19/USA/TX1/2020|EPI_ISL_411956; B4-1, hCoV-19/South Korea/KUMC03/2020|EPI_ISL_413513; B4-2, hCoV-19/USA/CA7/2020|EPI_ISL_411954.