

1 **SARS-CoV-2 detection in nasopharyngeal throat swabs by**
2 **metagenomics**

3

4 Le Van Tan¹, Nguyen Thi Thu Hong¹, Nghiem My Ngoc², Tran Tan Thanh¹, Vo Thanh
5 Lam², Lam Anh Nguyet¹, Le Nguyen Truc Nhu¹, Nguyen Thi Han Ny¹, Ngo Ngoc Quang
6 Minh³, Dinh Nguyen Huy Man², Vu Thi Ty Hang¹, Phan Nguyen Quoc Khanh¹, Tran Chanh
7 Xuan⁴, Nguyen Thanh Phong², Tran Nguyen Hoang Tu², Tran Tinh Hien^{1,5}, Le Manh
8 Hung², Nguyen Thanh Truong², Lam Minh Yen¹, Nguyen Thanh Dung², Guy Thwaites^{1,5},
9 Nguyen Van Vinh Chau², for OUCRU COVID-19 research group*

10 *Members of the Group are listed in the acknowledgements

11

12 ¹Oxford University Clinical Research Unit, Ho Chi Minh City, Vietnam

13 ²Hospital for Tropical Diseases, Ho Chi Minh City, Vietnam

14 ³Children's Hospital 1, Ho Chi Minh City, Vietnam

15 ⁴Cu Chi Hospital, Ho Chi Minh City, Vietnam

16 ⁵Centre for Tropical Medicine and Global Health, Nuffield Department of Medicine,
17 University of Oxford, Oxford, UK

18 **Correspondence:** Le Van Tan, email: tanlv@oucru.org

19 **Abstract:** 49

20 **Main text:** 1198

21 **Running title:** SARS-CoV-2 detection by metagenomics

22 **Key words:** COVID-19, SARS-CoV-2, coronaviruses, pandemic, Vietnam

23 **ABSTRACT**

24 Metagenomics could detect SARS-CoV-2 in all eight nasopharyngeal/throat swabs with
25 high/low viral loads, and rhinovirus in a co-infected patient. The sequenced viruses
26 belonged to lineage B1. Because metagenomics could detect novel pathogen and co-
27 infection, and generate sequence data for epidemiological investigation, it is an attractive
28 approach for infectious-disease diagnosis.

29 Metagenomics is a sensitive sequence-independence method for infectious disease diagnosis
30 and the discovery of novel pathogens [1]. The novel coronavirus namely severe acute
31 respiratory syndrome coronavirus 2 (SARS-CoV-2) is the cause of the ongoing coronavirus
32 disease 2019 (COVID-19) pandemic [2]. However, there have only been three studies
33 reporting the utility potential of metagenomics to detect SARS-CoV-2 directly from clinical
34 specimens, with a combined sample size of nine patients [3-5]. But none of these has been
35 conducted in resource-limited settings. In this area of the world, emerging infection however
36 is likely to emerge. Here we describe the application of metagenomics to detect SARS-CoV-
37 2 in RT-PCR positive nasopharyngeal throat swabs. In addition, using the obtained
38 sequence, we genetically characterize the viruses.

39 **THE STUDY**

40 Since the beginning of March, 2020 an observational study have been conducted at the
41 Hospital for Tropical Diseases (HTD) in Ho Chi Minh City, Vietnam and another one at one
42 of its two designated centres for receiving and treating COVI-19 patients from southern
43 Vietnam with a population of over 40 million (Figure 1). We enrolled patients with a
44 confirmed SARS-CoV-2 diagnosis admitted to the study settings within 48 hours. We
45 collected nasopharyngeal throat swabs (NTS), clinical and laboratory data, and travel and
46 contact history from each study participant. The collected NTS was stored at 4⁰C at the
47 study sites within four hours and was then transferred to the clinical laboratory of HTD for
48 analysis. SARS-CoV-2 detection was carried out using a WHO recommended real time RT-
49 PCR assays [6]. Assessment of co-infection with common respiratory viruses was carried
50 out using multiplex RT-PCR targeted at 15 different respiratory viruses [7]. The clinical
51 studies received approvals from the Institutional Review Board of the HTD and the Oxford

52 Tropical Research Ethics Committee of the University of Oxford. Study participants gave
53 their written informed consent.

54 The selected samples were individually analyzed with the inclusion of a molecular grade
55 water sample serving as a non-template control (NTC). Metagenomics was carried out as
56 previously described [8]. DNA libraries of individual samples and NTC were then
57 multiplexed using double unique indexes (i.e. each sample was differentiated by double
58 barcodes) and sequenced on an Illumina MiSeq platform using a 300-cycle MiSeq reagent
59 kit V3 (Illumina). Detection of SARS-CoV-2 and co-infection viruses in the obtained
60 sequence data was carried out using a combination of publically available metagenomics
61 pipelines namely IDseq (idseq.net) and DISCVR [9]. Reference based mapping approach
62 was applied to assemble SARS-CoV-2 genomes from the obtained sequences using
63 Geneious 11.0.3 (Biomatters, Auckland, New Zealand). SARS-CoV-2 lineage determination
64 and detections of nonsynonymous mutations were carried out using CoV-GLUE ([http://cov-
65 glue.cvr.gla.ac.uk](http://cov-glue.cvr.gla.ac.uk)), a publically available tool for SARS-CoV-2 sequence analysis (Figure
66 1).

67 As of March 19th, 2020, a total of 11 PCR confirmed SARS-CoV-2 patients were enrolled in
68 the clinical studies (Figure 1). As a pilot, we selected eight with a wide range of viral loads,
69 as reflected by real time Cycle threshold (Ct) values, for metagenomics analysis (Figure
70 2A). Information about demographics and clinical status of the eight included patients are
71 presented in Table 1. All were adults and two were asymptomatic carriers identified through
72 contact tracing approach implemented in Vietnam [10]. Three were cases of locally acquired
73 infection and five were imported cases, and one was co-infected with rhinovirus.

74 Information about duration of stay and clinical and laboratory findings are presented in
75 Table 1.
76 Metagenomics generated a total of 2–4 million reads per sample in 7/8 included NTS. In the
77 remaining sample, ¼ million reads were obtained (Table 2). SARS-CoV-2 were detected in
78 sequence data obtained from all eight RT-PCR positive NTS samples by both IDseq and
79 DISCVR, but not in the NTS sample. One patient presenting with respiratory infection was
80 co-infected with rhinovirus, which was also detected by metagenomics.
81 Results of reference-based mapping showed three consensus had genome coverage of
82 $\geq 70\%$, while the remaining five had coverage of $< 50\%$ (Table 2 and Supplementary Figure
83 1). Analysis of the obtained consensus showed all belong to lineage B1. A total of 11
84 nonsynonymous substitutions were detected in three of the eight obtained consensus
85 (Supplementary Table 1).

86 **CONCLUSIONS**

87 The emergence of SARS-CoV-2 emphasizes the continuous unprecedented threat posed by
88 emerging infectious diseases, especially those caused by novel viruses. The diagnosis of
89 respiratory diseases is highly challenging because the responsible pathogens are diverse. In
90 addition, the emergence of novel pathogens further challenges routine diagnosis. Indeed,
91 SARS-CoV-2 initially went undetected by PCR panels targeted at common respiratory
92 viruses [2]. New diagnostic approach is therefore urgently needed to address the ongoing
93 challenge posed by emerging infections.

94 Here, we demonstrated that when coupled with publically available bioinformatics tools,
95 metagenomics could detect SARS-CoV-2 in RT-PCR positive NTS samples with a wide
96 range of viral loads. The data suggests that metagenomics is a sensitive assay for SARS-

97 CoV-2 diagnosis and detection of co-infection as illustrated by the detection of rhinovirus,
98 in line with a recent report [4], important for clinical management. In addition to providing
99 diagnostic information, the obtained sequences also allows for genetic characterization, and
100 detection of genetic variations in the genomes of the pathogen under investigation. Indeed,
101 using the obtained sequences, we successfully identified that all the Vietnamese viruses
102 included for analysis belonged to lineage B1, which has been found worldwide [11]. In line
103 with a recent report [12], we identify several nonsynonymous substitutions in the obtained
104 genomes SARS-CoV-2. Further research is needed to ascribe the potential consequences
105 that SARS-CoV-2 evolution may have.

106 Currently, real time RT-PCR is used for screening of suspected cases of SARS-CoV-2
107 infection [6]. Compared with RT-PCR, metagenomics based on Illumina sequencing
108 technologies remains high cost and low throughput. However, these caveats could be
109 overcome by third generation sequencing technologies such as Oxford Nanopore [13],
110 which warrants further research.

111 The application of metagenomics for SARS-CoV-2 and respiratory diagnosis would be
112 highly relevant in the near future. This is because SARS-CoV-2 has spread globally, and
113 will likely soon become endemic worldwide. Indeed as of May 21st, 2020 nearly 5 million
114 cases have been reported globally. Notably, the vast majority of SARS-CoV-2 infections are
115 asymptomatic or mild, while COVID-19 patients present with signs/symptoms
116 undistinguished with respiratory diseases caused by other viruses [14, 15]. As such rapid
117 identification of the likely cause of hospitalized patients with respiratory infections is
118 essential for clinical management and outbreak response. Under this circumstance,
119 metagenomics is a preferable method because of its ability to detect both known and

120 unknown pathogens presenting in the tested specimens without the need of pathogen
121 specific PCR primers [1, 13].
122 Our study has some limitations. Only a small number of patients were included for analysis,
123 owing to the nature of a pilot in itself. However during the study period, there were only 14
124 SARS-CoV-2 confirmed cases reported in our setting, Ho Chi Minh City, Vietnam. As a
125 consequence, we were not able to properly assess the sensitivity and specificity of
126 metagenomics for the diagnosis of COVID-19.
127 In summary, we show that metagenomics is a sensitive assay for sequence-independent
128 detection of SARS-CoV-2 NTS samples. The ability of metagenomics to detect co-infection
129 and novel pathogens, and generate sequence data for molecular epidemiological
130 investigation makes it an attractive approach for infectious disease diagnosis.

131 **ACKNOWLEDGEMENTS**

132 This study was funded by the Wellcome Trust of Great Britain (106680/B/14/Z and
133 204904/Z/16/Z).

134 We are indebt to Ms Nguyen Thanh Ngoc, Ms Le Kim Thanh, and the OUCRU
135 IT/CTU/Laboratory Management departments for their support.

136

137 We thank the patients for their participations in this study, and the doctors and nurses at
138 HTD Cu Chi Hospital, who cared for the patients and provided the logistic support with the
139 study.

140 **OUCRU COVID-19 Research Group**

141 **Hospital for Tropical Diseases, Ho Chi Minh City, Vietnam:** Nguyen Van Vinh Chau,
142 Nguyen Thanh Dung, Le Manh Hung, Huynh Thi Loan, Nguyen Thanh Truong, Nguyen
143 Thanh Phong, Dinh Nguyen Huy Man, Nguyen Van Hao, Duong Bich Thuy, Nghiem My
144 Ngoc, Nguyen Phu Huong Lan, Pham Thi Ngoc Thoa, Tran Nguyen Phuong Thao, Tran Thi
145 Lan Phuong, Le Thi Tam Uyen, Tran Thi Thanh Tam, Bui Thi Ton That, Huynh Kim
146 Nhung, Ngo Tan Tai, Tran Nguyen Hoang Tu, Vo Trong Vuong, Dinh Thi Bich Ty, Le Thi
147 Dung, Thai Lam Uyen, Nguyen Thi My Tien, Ho Thi Thu Thao, Nguyen Ngoc Thao,
148 Huynh Ngoc Thien Vuong, Pham Ngoc Phuong Thao, Phan Minh Phuong

149 **Oxford University Clinical Research Unit, Ho Chi Minh City, Vietnam:** Dong Thi Hoai
150 Tam, Evelyne Kestelyn, Donovan Joseph, Ronald Geskus, Guy Thwaites, H. Rogier van
151 Doorn, Huynh Le Anh Huy, Huynh Ngan Ha, Huynh Xuan Yen, Jennifer Van Nuil, Jeremy
152 Day, Joseph Donovan, Katrina Lawson, Lam Anh Nguyet, Lam Minh Yen, Le Nguyen Truc
153 Nhu, Le Thanh Hoang Nhat, Le Van Tan, Sonia Lewycka Odette, Louise Thwaites, Maia
154 Rabaa, Marc Choisy, Mary Chambers, Motiur Rahman, Ngo Thi Hoa, Nguyen Thanh Thuy
155 Nhien, Nguyen Thi Han Ny, Nguyen Thi Kim Tuyen, Nguyen Thi Phuong Dung, Nguyen
156 Thi Thu Hong, Nguyen Xuan Truong, Phan Nguyen Quoc Khanh, Phung Le Kim Yen,
157 Sophie Yacoub, Thomas Kesteman, Nguyen Thuy Thuong Thuong, Tran Tan Thanh, Tran
158 Tinh Hien, Vu Thi Ty Hang

159 **ABOUT THE AUTHOR**

160 Dr Le Van Tan is head of Emeging Infections at Oxford Univeristy Clinical Resarh Unit.

161 His research interest includes novel diagnosis and emerging infections.

162

163 **REFERENCES**

- 164 1. Chiu, C.Y. and S.A. Miller, *Clinical metagenomics*. Nat Rev Genet, 2019. **20**(6): p.
165 341-355.
- 166 2. Zhu, N., D. Zhang, W. Wang, X. Li, B. Yang, J. Song, X. Zhao, B. Huang, W. Shi,
167 R. Lu, P. Niu, F. Zhan, X. Ma, D. Wang, W. Xu, G. Wu, G.F. Gao, W. Tan, I. China

- 168 Novel Coronavirus, and T. Research, *A Novel Coronavirus from Patients with*
169 *Pneumonia in China, 2019*. N Engl J Med, 2020. **382**(8): p. 727-733.
- 170 3. Chen, L., W. Liu, Q. Zhang, K. Xu, G. Ye, W. Wu, Z. Sun, F. Liu, K. Wu, B. Zhong,
171 Y. Mei, W. Zhang, Y. Chen, Y. Li, M. Shi, K. Lan, and Y. Liu, *RNA based mNGS*
172 *approach identifies a novel human coronavirus from two individual pneumonia*
173 *cases in 2019 Wuhan outbreak*. Emerg Microbes Infect, 2020. **9**(1): p. 313-319.
- 174 4. Peddu, V., R.C. Shean, H. Xie, L. Shrestha, G.A. Perchetti, S.S. Minot, P.
175 Roychoudhury, M.L. Huang, A. Nalla, S.B. Reddy, Q. Phung, A. Reinhardt, K.R.
176 Jerome, and A.L. Greninger, *Metagenomic analysis reveals clinical SARS-CoV-2*
177 *infection and bacterial or viral superinfection and colonization*. Clin Chem, 2020.
- 178 5. Ai, J.W., Y. Zhang, H.C. Zhang, T. Xu, and W.H. Zhang, *Era of molecular*
179 *diagnosis for pathogen identification of unexplained pneumonia, lessons to be*
180 *learned*. Emerg Microbes Infect, 2020. **9**(1): p. 597-600.
- 181 6. Corman, V.M., O. Landt, M. Kaiser, R. Molenkamp, A. Meijer, D.K. Chu, T.
182 Bleicker, S. Brunink, J. Schneider, M.L. Schmidt, D.G. Mulders, B.L. Haagmans, B.
183 van der Veer, S. van den Brink, L. Wijsman, G. Goderski, J.L. Romette, J. Ellis, M.
184 Zambon, M. Peiris, H. Goossens, C. Reusken, M.P. Koopmans, and C. Drosten,
185 *Detection of 2019 novel coronavirus (2019-nCoV) by real-time RT-PCR*. Euro
186 *Surveill*, 2020. **25**(3).
- 187 7. Jansen, R.R., J. Schinkel, S. Koekkoek, D. Pajkrt, M. Beld, M.D. de Jong, and R.
188 Molenkamp, *Development and evaluation of a four-tube real time multiplex PCR*
189 *assay covering fourteen respiratory viruses, and comparison to its corresponding*
190 *single target counterparts*. J Clin Virol, 2011. **51**(3): p. 179-85.
- 191 8. Anh, N.T., N.T.T. Hong, L.N.T. Nhu, T.T. Thanh, C.Y. Lau, D. Limmathurotsakul,
192 X. Deng, M. Rahman, N.V.V. Chau, H.R. van Doorn, G. Thwaites, E. Delwart, and
193 L.V. Tan, *Viruses in Vietnamese Patients Presenting with Community-Acquired*
194 *Sepsis of Unknown Cause*. J Clin Microbiol, 2019. **57**(9).
- 195 9. Maabar, M., A.J. Davison, M. Vucak, F. Thorburn, P.R. Murcia, R. Gunson, M.
196 Palmarini, and J. Hughes, *DisCVR: Rapid viral diagnosis from high-throughput*
197 *sequencing data*. Virus Evol, 2019. **5**(2): p. vez033.
- 198 10. Dinh, L., P. Dinh, P.D.M. Nguyen, D.H.N. Nguyen, and T. Hoang, *Vietnam's*
199 *response to COVID-19: Prompt and proactive actions*. J Travel Med, 2020.
- 200 11. Rambaut, A., E.C. Holmes, V. Hill, Á. O'Toole, J.T. McCrone, C. Ruis, L. du
201 Plessis, and O.G. Pybus, 2020.
- 202 12. van Dorp, L., M. Acмна, D. Richarch, P.L. Shaw, E.C. Fprd, L. Ormond, J.C. Owen,
203 J. Pang, C.S.C. Tan, A.T.F. Boshier, T.A. Ortiz, and F. Balloux, *Emergence of*
204 *genomic diversity and recurrent mutations in SARS-CoV-2*. Infection, Genetics and
205 *Evolution*, 2020. **in press**.
- 206 13. Yang, L., G. Haidar, H. Zia, R. Nettles, S. Qin, X. Wang, F. Shah, S.F. Rapport, T.
207 Charalampous, B. Methe, A. Fitch, A. Morris, B.J. McVerry, J. O'Grady, and G.D.
208 Kitsios, *Metagenomic identification of severe pneumonia pathogens in mechanically-*
209 *ventilated patients: a feasibility and clinical validity study*. Respir Res, 2019. **20**(1):
210 p. 265.
- 211 14. Goyal, P., J.J. Choi, L.C. Pinheiro, E.J. Schenck, R. Chen, A. Jabri, M.J. Satlin, T.R.
212 Champion, Jr., M. Nahid, J.B. Ringel, K.L. Hoffman, M.N. Alshak, H.A. Li, G.T.
213 Wehmeyer, M. Rajan, E. Reshetnyak, N. Hupert, E.M. Horn, F.J. Martinez, R.M.

- 214 Gulick, and M.M. Safford, *Clinical Characteristics of Covid-19 in New York City*. N
215 Engl J Med, 2020.
- 216 15. Grasselli, G., A. Zangrillo, A. Zanella, M. Antonelli, L. Cabrini, A. Castelli, D.
217 Cereda, A. Coluccello, G. Foti, R. Fumagalli, G. Iotti, N. Latronico, L. Lorini, S.
218 Merler, G. Natalini, A. Piatti, M.V. Ranieri, A.M. Scandroglio, E. Storti, M.
219 Cecconi, A. Pesenti, and C.-L.I. Network, *Baseline Characteristics and Outcomes of*
220 *1591 Patients Infected With SARS-CoV-2 Admitted to ICUs of the Lombardy Region,*
221 *Italy*. JAMA, 2020.

Table 1: Demographics, clinical and real time RT-PCR data of the study participants

	Patient 1	Patient 2	Patient 3	Patient 4	Patient 5	Patient 6	Patient 7	Patient 8
Age range	30's	40's	20's	30's	20's	20's	40's	10's
Gender	Male	Male	Male	Male	Female	Female	Female	Male
Arriving in Vietnam from abroad (Yes/No)	Yes	Yes	No	No	No	Yes	Yes	Yes
Locally acquired infection (Yes/No)	No	No	Yes	Yes	Yes	No	No	No
Days from confirmed diagnosis to enrolment	2	1	3	2	3	2	1	0
Days from admission to enrolment	2	1	2	1	0	1	2	0
Duration of stay	15	NA	17	18	15	12	13	17
Symptomatic (S)/asymptomatic (A)	S	S	S	A	S	A	S	S
Laboratory results								
White-cell count (×103per µl)	4.23	6.89	3.96	NA	6.85	4.83	3.46	8.27
Lymphocyte counts (×103per µl)	0.8	0.54	1.08	NA	2.94	2.50	0.88	2.40
Hemoglobin ((g/dl)	13.6	14.6	16.8	NA	15.7	15.00	11.60	15.60
Hematocrit (%)	42.4	43.4	41.6	NA	37.4	35.78	28.48	38.09
Platelet count (per µl)	140	235	187	NA	414	321.00	178.00	330.00
Glucose (mg/dl)	125.7	112	85	NA	NA	NA	76.80	98.50
Creatinine (mg/dl)	0.96	1.28	1.2	NA	NA	NA	0.99	1.19
Aspartate aminotransferase (U/liter)	24	17	22	NA	NA	NA	17.40	23.30
Alanine aminotransferase (U/liter)	23	16	24	NA	NA	NA	17.40	25.30
Clinical signs/symptoms (Yes/No)								
Fever	Y	Y	Y	N	N	N	N	N
Cough	N	Y	N	N	N	N	Y	N
Rhinorrhea	N	N	Y	N	Y	N	N	Y
Fatigue	N	Y	N	N	N	N	N	N
Diarrhea	N	N	N	N	N	N	N	Y
Sore throat	N	N	Y	N	N	N	Y	N
Muscle pain	N	Y	N	N	N	N	Y	N
Headache	N	Y	N	N	N	N	N	N
Abdominal pain	N	N	N	N	N	N	N	N
Lost sense of smell	N	N	N	N	Y	N	N	N

Table 2: Results of mNGS and lineage assignment SARS-CoV-2 sequences

Patient number	Ct values	Total single reads	No of SARS-CoV-2 reads	% SARS-CoV-2 genome coverage	SARS-CoV-2 lineage
1	26.52	3,182,758	84	16	B1
2	21.47	4,218,464	6930	70	B1
3	27.06	2,735,464	573	36	B1
4	32.09	1,902,512	68	12	B1
5	24.81	245,818	14	3	B1
6	25.68	3,524,972	995	48	B1
7	24.56	2,440,326	16564	87	B1
8	24.13	3,253,308	9095	80	B1

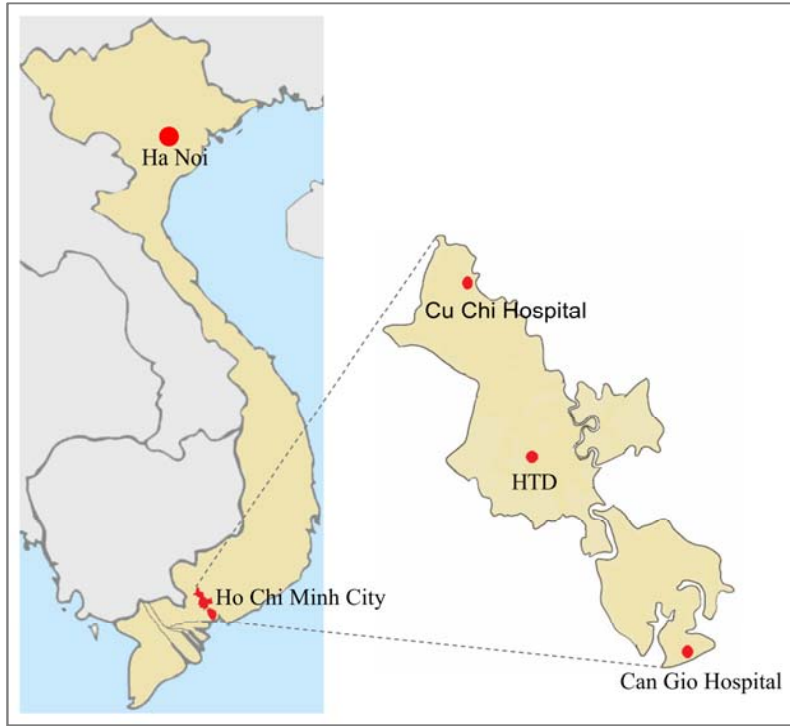
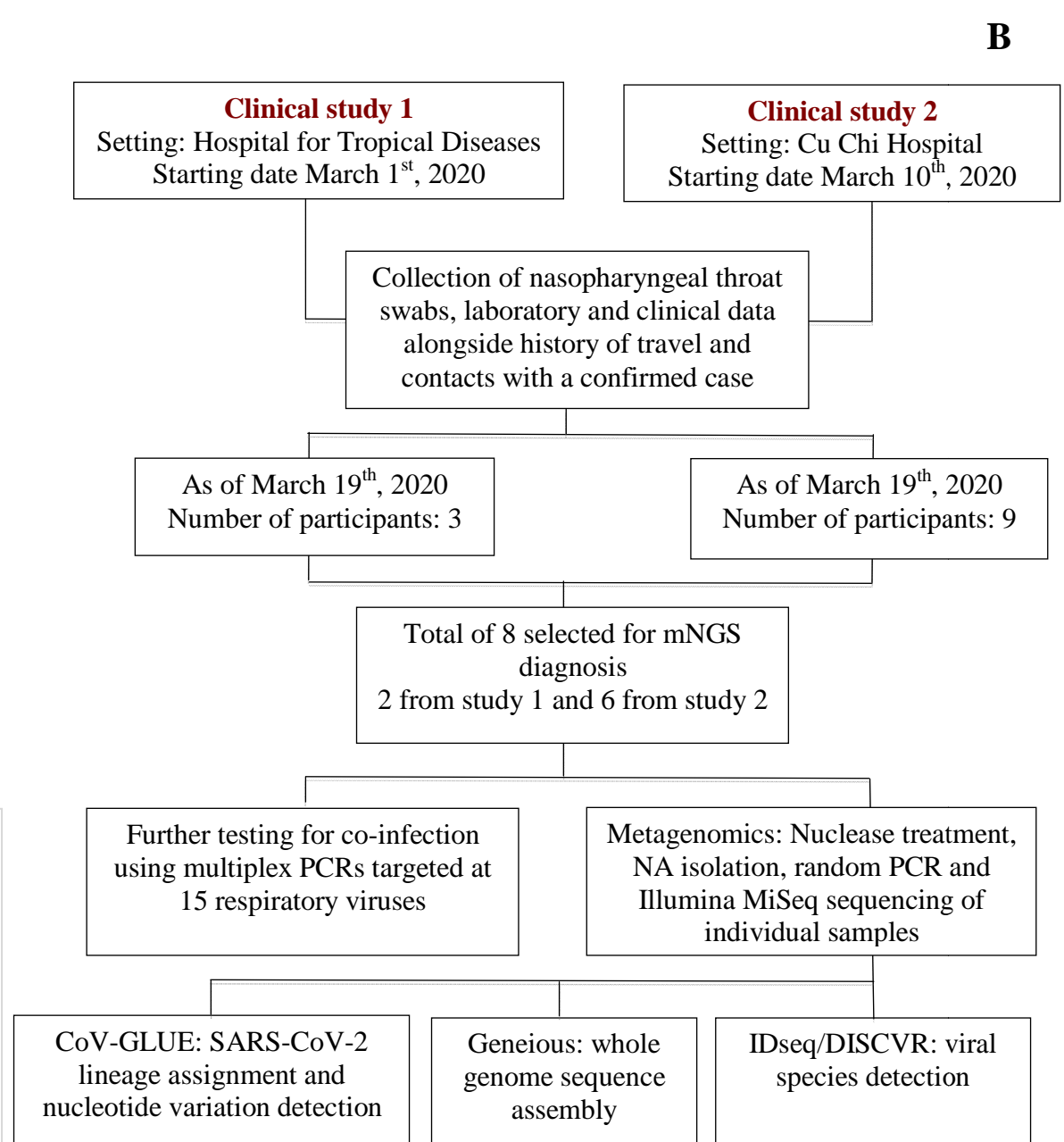


Figure 1: Flowchart showing the study settings and the clinical studies. **(A)** Map showing the location of the Hospital for Tropical Diseases (HTD) main campus and its designated COVID-19 centres in Cu Chi where the clinical study was conducted, and Can Gio, **(B)** Flowchart illustrating the patient enrolment, mNGS analysis and post run analysis of the obtained sequences

Note to Figure 1:

Maps were obtained from <https://mapchart.net/>.



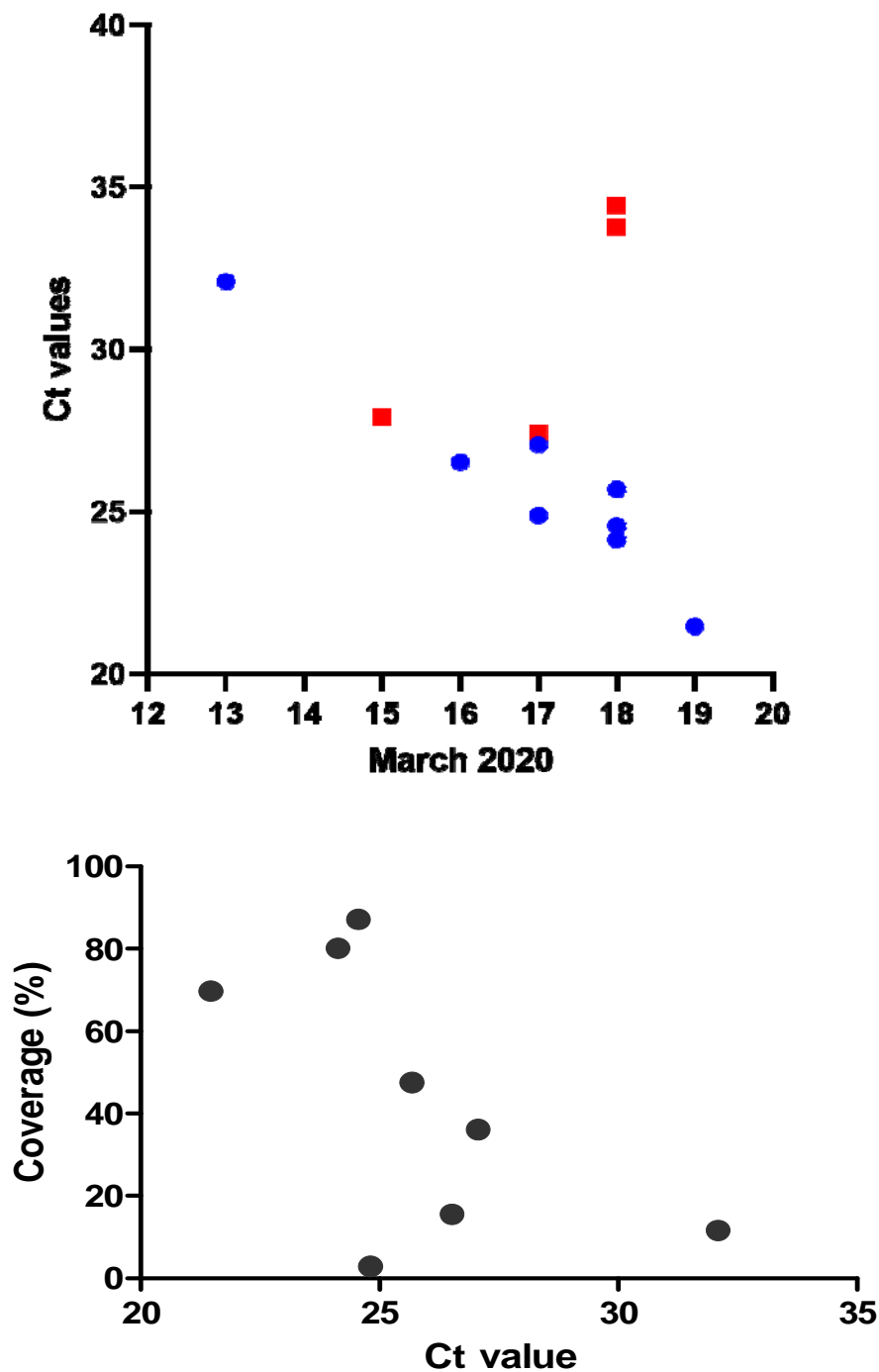


Figure 2: Distribution of Ct values of nasopharyngeal throat swabs of the study participant (A), and the association between Ct values and genome coverage of SARS-CoV-2 generated by mNGS (B)

Note to Figure 2A: Blue dot represent for samples selected for mNGS while red squares represent for samples not selected for mNGS. Numbers on the X axis represent for calendar days of March 2020

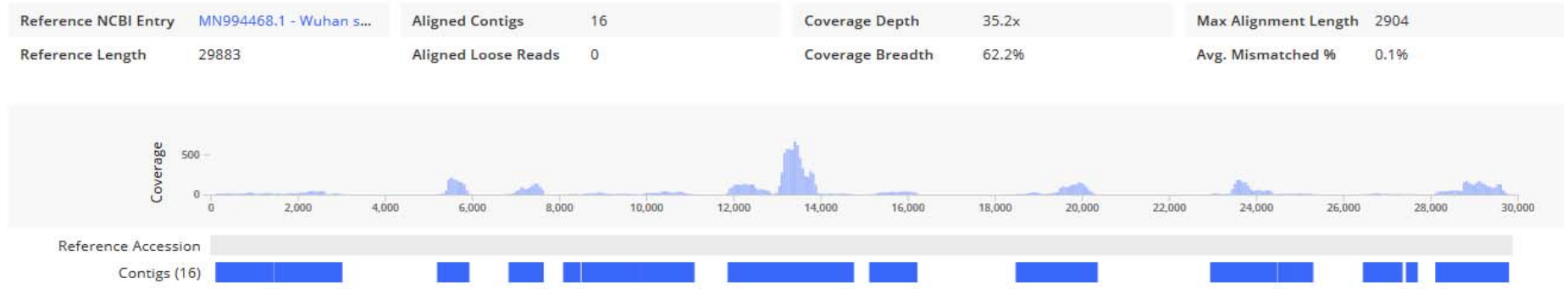
Supplementary Table 1: list of non synonymous substitution detected in eight consensuses of the present study

	Nucleotide variation*	Coding Region	Amino acid change	Detected in GenBank
Patient 6	17104C>T	nsp13	H290Y	Yes
Patient 7	14407C>T	nsp12	P323L	Yes
	23402G>A	S	D614G	Yes
	28881G>A	N	R203K	Yes
	28882G>A	N	R203K	Yes
	28883G>C	N	G204R	Yes
Patient 8	14407C>T	nsp12	P323L	Yes
	23402G>A	S	D614G	Yes
	28881G>A	N	R203K	Yes
	28882G>A	N	R203K	Yes
	28883G>C	N	G204R	Yes

Note to supplementary Figure 1: *compared with reference strain (GenBank accession number NC_045512.2)

Betacoronavirus Coverage - Wuhan seafood market pneumonia virus ⓘ
MN994468.1 - Wuhan seafood market pneumonia virus isolate 201... ▾
4 viewable accessions

[View read-level visualization >](#)



Supplementary Figure 1: A screen shot showing evidence of SARS-CoV-2 detection in metagenomics data using IDseq pipeline