Genomic surveillance of SARS-CoV-2 in Thailand reveals mixed imported populations, a local lineage expansion and a virus with truncated ORF7a

Khajohn Joonlasak [a,b,1], Elizabeth M Batty [a,c,d,1], Theerarat Kochakarn [a,e,f,1], Bhakbhoom Panthan [a,g,1], Krittikorn Kümpornsin [a,h], Poramate Jiaranai [a,g], Arporn Wangwiwatsin [a,i], Angkana Huang [a,b], Namfon Kotanan [a,e], Peera Jaru-Ampornpan [j], Wudtichai Manasatienkij [a,b], Treewat Watthanachockchai [k], Kingkan Rakmanee [k], Anthony R. Jones [a,b], Stefan Fernandez [a,b], Insee Sensorn [a,g], Somnuek Sungkanuparph [l], Ekawat Pasomsub [a,k], Chonticha Klungthong [a,b], Thanat Chookajorn [a,e,*], Wasun Chantratita [a,g,*]

[a] COVID-19 Network Investigations Alliance (CONI), Bangkok, Thailand
[b] Department of Virology, Armed Forces Research Institute of Medical Sciences, Bangkok, Thailand
[c] Mahidol-Oxford Tropical Medicine Research Unit, Faculty of Tropical Medicine, Mahidol University, Bangkok, Thailand
[d] Centre for Tropical Medicine and Global Health, Nuffield Department of Medicine, University of Oxford, Oxford, United Kingdom
[e] Genomics and Evolutionary Medicine Unit (GEM), Center of Excellence in Malaria Research, Faculty of Tropical Medicine, Mahidol University, Bangkok, Thailand
[f] Faculty of Medicine Ramathibodi Hospital, Mahidol University Bangkok, Thailand
[g] Center for Medical Genomics, Faculty of Medicine Ramathibodi Hospital, Mahidol University, Bangkok, Thailand
[h] Wellcome Sanger Institute, Hinxton, Cambridgeshire, United Kingdom
[i] Department of Biology, Faculty of Science, Khon Kaen University, Khon Kaen, Thailand
[j] Virology and Cell Technology Research Team, National Center for Genetic Engineering and Biotechnology (BIOTEC), National Science and Technology Development Agency (NSTDA), Pathum Thani, Thailand
[k] Division of Virology, Department of Pathology, Faculty of Medicine Ramathibodi Hospital, Mahidol University, Bangkok, Thailand
[l] Chakri Naruebodindra Medical Institute, Faculty of Medicine Ramathibodi Hospital, Mahidol University, Samut Prakan, Thailand

ARTICLE INFO

ABSTRACT

Coronavirus Disease 2019 (COVID-19) is a global public health threat. Genomic surveillance of SARS-CoV-2 was implemented in March of 2020 at a major diagnostic hub in Bangkok, Thailand. Several virus lineages supposedly originated in many countries were found, and a Thai-specific lineage, designated A/Thai-1, has expanded to be predominant in Thailand. A virus sample in the SARS-CoV-2 A/Thai-1 lineage contains a frame-shift deletion at ORF7a, encoding a putative host antagonizing factor of the virus.

1. Introduction

Coronavirus Disease 2019 (COVID-19) has reached the status of global pandemic. Genomic surveillance of its etiological virus, SARS-CoV-2, plays an important role in epidemiological investigations and transmission control strategies (Gudbjartsson et al., 2020). Genetic variation data of the virus could reveal transmission chains between infected individuals and could even map the connection between outbreak cohorts. Thailand has suffered from the spread of COVID-19 with the total number of confirmed cases over 3000 and with more

than 120,000 individuals screened as of May 2020. Since January 2020, when both imported and locally-transmitted COVID-19 cases were reported in Thailand, the country implemented several measures to combat COVID-19 on a national scale (Okada et al., 2020; Pongpirul et al., 2020; Hinjoy et al., 2020).

Genomic surveillance could be a powerful tool in the implementation of the national COVID-19 control strategy in Thailand. ARTIC multiplex tiling PCR allows whole-genome sequencing with minuscule amount of material by generating genome-wide overlapping amplicons, which has led to its success during the Zika virus outbreak investigation in Brazil
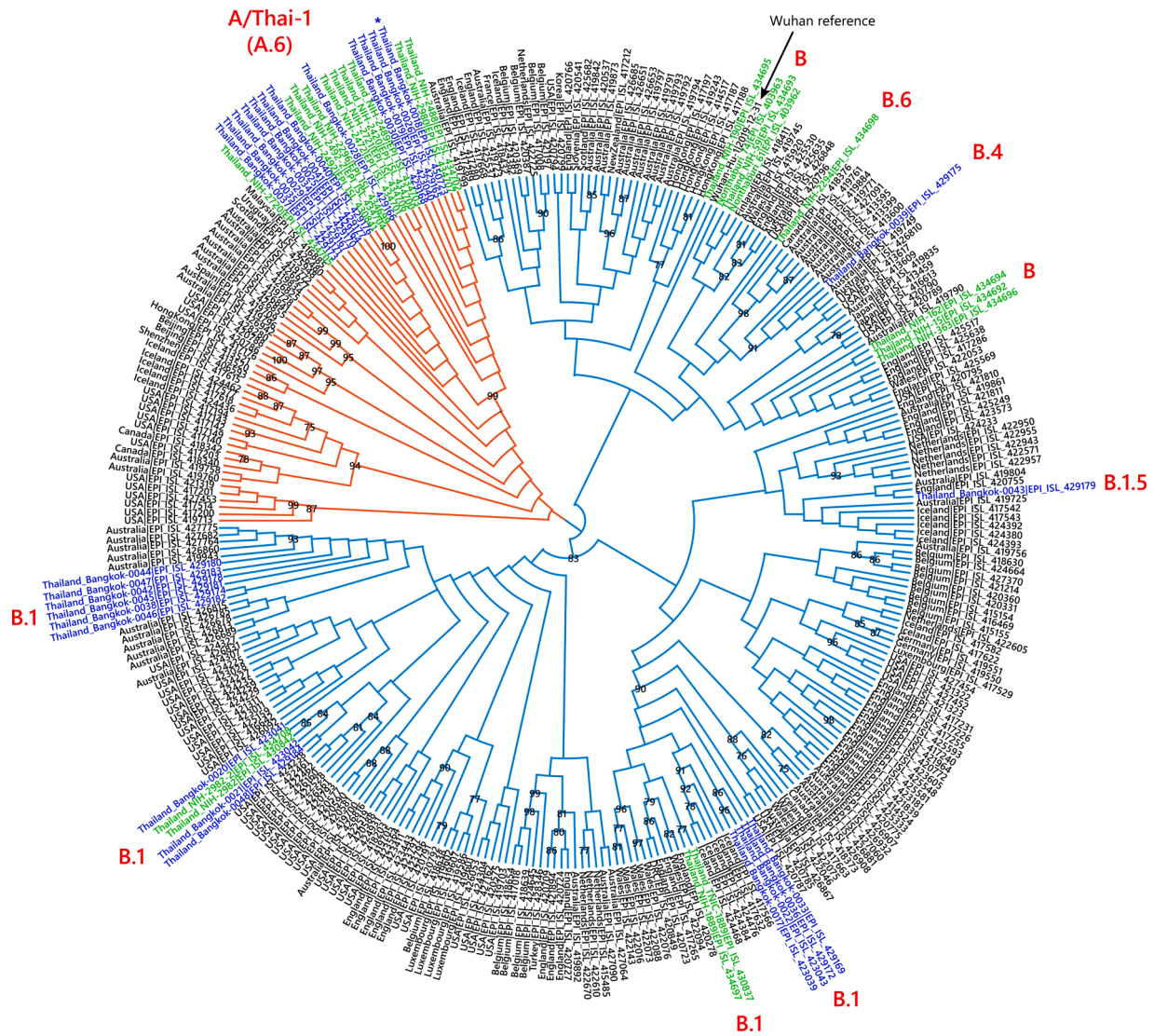
Fig. 1. Cladogram based on a maximum-likelihood tree showing Thai populations of SARS-CoV-2. The cladogram represents a matching maximum-likelihood tree (1,000 bootstrap replicates) shown in Supplementary Fig. 1. Genomes generated in this study are labelled in blue with remaining lineage representatives labelled in black. Bootstrap values (≥ 75) are shown at the nodes. Thai virus genomes independently generated by other groups and deposited in GISAID are labelled in green. Genomes collected from January 2020 are similar to that of reference Wuhan-Hu-1. Our data showed at least four independently lineages, and two additional events of lineage B from January 2020 and lineage B.6 (shown in green), could be recognized as potential introduction events. The A and B lineages are coloured in orange and cyan, respectively. * indicates the sample with the ORF7a deletion.

(Quick et al., 2017; Giovanetti et al., 2020). Using leftover RNA samples from a standard RT-PCR diagnosis, the genomic information of SARS-CoV-2 can be decoded in less than a week. The data presented here provide an insight into the genetic repertoire, origins and viral lineages of SARS-CoV-2 in Thailand. The information is particularly important given the multiple introduction events into the country and the local expansion of the Thai-specific SARS-CoV-2 lineages.

2. Genomic surveillance of SARS-CoV-2 populations in Thailand

We sequenced 27 anonymized RT-qPCR positive virus transport media samples containing nasopharyngeal/oropharyngeal swabs from Ramathibodi Hospital in Bangkok from March 13, 2020 to March 28, 2020 (Supplementary Table 1) [EC approval number: MURA2020/676]. The hospital acted as one of the major diagnostic hubs for COVID-19 in Bangkok during the study period. Enrichment and amplification steps were done according to the ARTIC Network protocol with ARTIC primer version 2 (Quick, 2020). The libraries were prepared using KAPA HyperPrep and KAPA Library Amplification kits and subsequently

sequenced using with a MiSeq Reagent Kit v2 according to the manufactures' protocols. Variant calling was performed using the ncov2019-artic-nf pipeline (https://github.com/connor-lab/ncov2019-artic-nf) using the default Illumina parameters (variants called by iVar v1.11 with minimum frequency threshold 0.75 and minimum depth 10) (Grubaugh et al., 2019). Consensus sequences were used to construct the maximum-likelihood and Bayesian phylogenetic trees with recommended representatives from various lineages worldwide utilizing IQ-TREE 2.0 and BEAST v1.10.4, respectively (Supplementary Table 2) (Drummond et al., 2012; Minh et al., 2020; Rambaut et al., 2020a). Interestingly, Thailand appears to have had multiple introduction events of SARS-CoV-2 into the country, as evidenced by at least six separate clusters in the maximum-likelihood tree (Fig. 1 and Supplementary Fig. 1) (Hadfield et al., 2018). Based on a Pangolin classification system (Database version 27 April 2020), they are grouped into A (12 samples), B.1 (13 samples), B.1.5 (1 sample) and B.4 (1 sample) lineages (Fig. 1, Supplementary Fig. 1 and Supplementary Table 1) (Rambaut et al., 2020b; O'Toole, 2020). The relationship and the origin of these lineages were described in (Rambaut et al., 2020a). Considering the origins and
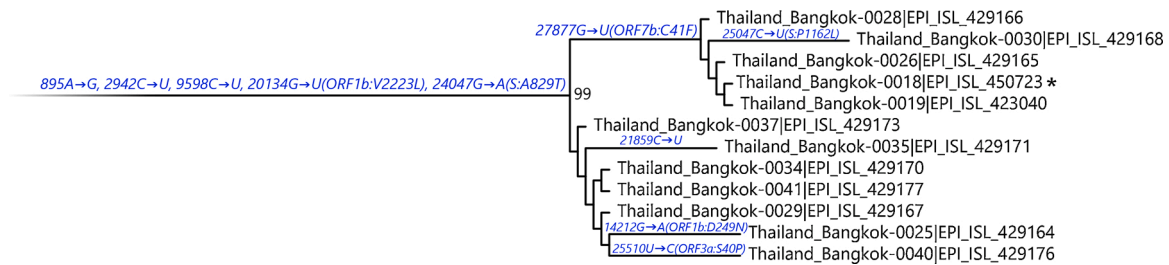
Fig. 2. Maximum-likelihood tree containing A/Thai-1 subset. For visualization of A/Thai-1 lineage, the A/Thai-1 genomes from Fig. 1 generated in this study are presented. The lineage defining node has a bootstrap value of 99 (1,000 replicates) as shown in Supplementary Fig. 1. Mutations that define each branch point are shown. 20,134G→U in Thailand_Bangkok-0030 could not be called. * indicates the sample with the ORF7a deletion.

Table 1
Mutations present in A/Thai-1 in comparison with Wuhan-Hu-1.

| Position	Ref	Alt	Syn/Non-syn	Gene	Samples	Note
895	A	G	S	ORF1a	0018, 0019, 0025, 0026, 0028, 0029, 0030, 0034, 0035, 0037, 0040, 0041	Unique to A/Thai-1
2942	C	U	S	ORF1a	0018, 0019, 0025, 0026, 0028, 0029, 0030, 0034, 0035, 0037, 0040, 0041	Unique to A/Thai-1
4390	G	U	NS	ORF1a	0018, 0019, 0025, 0026, 0028, 0029, 0030, 0034, 0035, 0037, 0040, 0041	Shared with Malaysia/MKAK-CL-2020-2096 (A lineage)
8782	C	U	S	ORF1a	0018, 0019, 0025, 0026, 0028, 0029, 0030, 0034, 0035, 0037, 0040, 0041	Shared among ancestral A lineages
9598	C	U	S	ORF1a	0018, 0019, 0025, 0026, 0028, 0029, 0030, 0034, 0035, 0037, 0040, 0041	Shared with Scotland/EDB2081 (B.1 lineage)
14212	G	A	NS	ORF1b	0025	–
20134	G	U	NS	ORF1b	0018, 0019, 0025, 0026, 0028, 0029, 0034, 0035, 0037, 0040, 0041	Cannot be called in 0030; Shared with Netherlands/NA_300 and USA/UN-UW-5172 (B.1 lineage)
21859	C	U	S	S	0035	–
24047	G	A	NS	S	0018, 0019, 0025, 0026, 0028, 0029, 0030, 0034, 0035, 0037, 0040, 0041	Unique to A/Thai-1
25047	C	U	NS	S	0030	–
25510	U	C	NS	ORF3a	0040	–
27694-27697	TTTC	–	Frame-shift deletion	ORF7a	0018	–
27877	G	U	NS	ORF7b	0018, 0019, 0026, 0028, 0030	Shared with Taiwan/170 and USA/UT-00514 (B.1 lineage)
28144	U	C	NS	ORF8	0018, 0019, 0025, 0026, 0028, 0029, 0030, 0034, 0035, 0037, 0040, 0041	Shared among ancestral A lineages
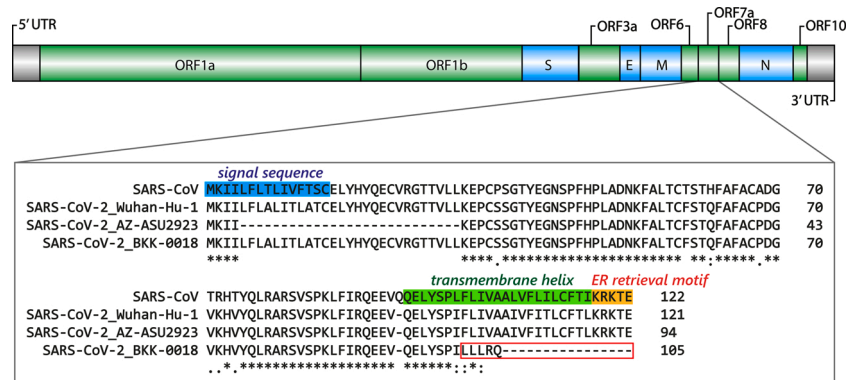
Fig. 3. Diagram depicting Thailand/Bangkok-0018 frame-shift deletion in ORF7a. The upper diagram shows the gene organization of the SARS-CoV-2 genome. The functional domains of ORF7a are indicated based on SARS-CoV. Sequences from SARS-CoV, the SARS-CoV-2 Wuhan-Hu-1 reference, the Arizona sample with an 81-nucleotide deletion (EPI_ISL_424669) and Thailand/Bangkok-0018 are aligned to demonstrate the altered region in red box.

lineage branches, these SARS-CoV-2 lineages are likely to have recent ancestors outside Thailand. For example, six of the samples in the B.1 lineage in our collection are grouped tightly with virus samples commonly found in the United States of America and Europe and collected during the same period (March 2020) as visualized by Nextstrain Timetree (Supplementary Fig. 2). The constructed Bayesian tree displayed the similar structure to that of the maximum-likelihood

approach (Supplementary Fig. 3).

It is worth noting the local expansion of a putative Thai specific lineage (Fig. 1 and Fig. 2). This cluster of viruses has passed the criteria of a novel lineage as follows: (a) exhibits two shared nucleotide differences from the ancestral lineage [cut point ≥ 1], (b) contains 11 genomes with > 95 % of the genome sequenced [cut point ≥ 5], (c) exhibits one shared nucleotide change (27,877G→U) among the

ongoing transmission groups from March 2020 [cut point ≥ 1], and (d) has 99 % bootstrap value for the lineage defining node [cut point > 70 %] (Rambaut et al., 2020b). This lineage, designated A/Thai-1 (Fig. 1 and 2), descended from the original A lineage (based on the maximum-likelihood based classification system), which was first reported in China before expanding into various countries in Asia, Europe, North America, South America and Australia. This A/Thai-1 branch (designated A.6 in the Pangolin system (Rambaut et al., 2020a)) is separated from the rest of the original A lineage and subgroups. Upon visual inspection in Nextstrain, only one Malaysian sample (MKAK-CL-2020-5096) is the closest to A/Thai-1, but only with 63 % bootstrap value by the maximum-likelihood approach and one shared lineage-specific nucleotide substitution (4,390G→U) (Supplementary Figure 4). Non-synonymous mutations unique to A/Thai-1 are 20, 134G→U (ORF1b) and 24,047G→A (Spike protein) (the full mutation list is shown in Table 1). Among the changes, 20,134G→U mutation has been independently found in two samples in lineage B.1 from the Netherlands and USA. It remains to be determined with a larger sample size whether this is the result of convergent evolution or genetic recombination. This pattern of homoplasy was also hypothetically linked to putative RNA editing (Simmonds, 2020).

Thailand/Bangkok-0018, a sample in the A/Thai-1 lineage, contains a 4-nt frame-shift deletion at position 27,694-27,697, causing a premature truncation in ORF7a, which now contains five altered amino acid residues and loses the 16 original C-terminal residues (Fig. 3). The deletion was confirmed by Sanger sequencing twice using two independent RT-PCR reactions (Supplementary Figure 5). The frame-shift mutation alters approximately one-sixth (21/121 residues) of the ORF7a protein. Based on protein homology to SARS-CoV, the missing region corresponds to a transmembrane helix and an ER retrieval motif, required for antagonizing a host antiviral factor (Taylor et al., 2015; Fielding et al., 2004). One sample from Arizona, USA also contains an 81-nt in-frame deletion in the ORF7a gene (Holland et al., 2020). So far, only one sample in A/Thai-1 appears to have this frame-shift deletion. It is tempting to speculate on the relationship between ORF7a deletions and virus attenuation. However, further investigations by laboratory-based functional experiments are needed before reaching any conclusion on their biological and clinical significance.

When the analysis was extended to an additional set of 22 genomes, independently deposited in GISAID by the Thai Red Cross and the Thai National Institute of Health, samples collected in January 2020 are grouped closely with the B lineage from China including the Wuhan-Hu-1 reference (Supplementary Table 3). The genetic repertoires from this additional collection also support the notion of multiple virus lineages introduced into Thailand. A/Thai-1 was the largest lineage in Thailand during the period of March 2020, with the total of 22 virus samples designated to A/Thai-1 (12 from our sequencing work and 10 from other independent genomic sequencing projects) from the total of 49 genomes available.

3. Implication of the findings

Genomic surveillance is likely to be pivotal in the identification and the elimination of transmission cohorts and chains (Kumporsin et al., 2019; Gardy and Loman, 2018). The genetic composition presented here suggests the necessity for screening and monitoring international travelers during the period of COVID-19 pandemic. The local expansion of A/Thai-1 strongly indicates a series of local transmission events, allowing an evolutionary branch unique to Thailand. This lineage needs to be investigated further for its compatibility to diagnosis and vaccine tools under development.

CRediT authorship contribution statement

Khajohn Joonlasak: Investigation. **Elizabeth M Batty:** Formal analysis, Software. **Theerarat Kochakarn:** Formal analysis, Software. **Bhakbhoom Panthan:** Investigation. **Krittikorn Kümpornsin:** Investigation, Methodology, Writing - review & editing. **Poramate Jiaranai:** Investigation. **Arporn Wangwiwatsin:** Investigation, Methodology. **Angkana Huang:** Formal analysis, Software. **Namfon Kotanan:** Investigation. **Peera Jaru-Ampornpan:** Formal analysis. **Wudtichai Manasatienkij:** Formal analysis. **Treewat Watthanachockchai:** Resources. **Kingkan Rakmanee:** Resources. **Anthony R. Jones:** Supervision, Writing - review & editing. **Stefan Fernandez:** Supervision, Writing - review & editing. **Insee Sensorn:** Project administration. **Somnuek Sungkanuparph:** Resources, Writing - review & editing. **Ekawat Pasomsub:** Resources, Formal analysis. **Chonticha Klungthong:** Supervision, Formal analysis, Writing - review & editing. **Thanat Chookajorn:** Formal analysis, Supervision, Project administration, Writing - original draft, Funding acquisition, Writing - review & editing. **Wasun Chantratita:** Formal analysis, Supervision, Project administration, Funding acquisition, Writing - review & editing.

Declaration of Competing Interest

None.

Acknowledgements

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:https://doi.org/10.1016/j.virusres.2020.198233.

References

Drummond, A.J., et al., 2012. Bayesian phylogenetics with BEAUti and the BEAST 1.7. Mol. Biol. Evol. 29 (8), 1969–1973.

Fielding, B.C., et al., 2004. Characterization of a unique group-specific protein (U122) of the severe acute respiratory syndrome coronavirus. J. Virol. 78 (14), 7311–7318.

Gardy, J.L., Loman, N.J., 2018. Towards a genomics-informed, real-time, global pathogen surveillance system. Nat. Rev. Genet. 19 (1), 9–20.

Giovanetti, M., et al., 2020. Genomic and epidemiological surveillance of zika virus in the Amazon region. Cell Rep. 30 (7), 2275–2283 e7.

Grubaugh, N.D., et al., 2019. An amplicon-based sequencing framework for accurately measuring intrahost virus diversity using PrimalSeq and iVar. Genome Biol. 20 (1), 8.

Gudbjartsson, D.F., et al., 2020. Spread of SARS-CoV-2 in the icelandic population. N. Engl. J. Med.

Hadfield, J., et al., 2018. Nextstrain: real-time tracking of pathogen evolution. Bioinformatics 34 (23), 4121–4123.

Hinjoy, S., et al., 2020. A self-assessment of the Thai Department of Disease Control's communication for international response at early phase to the COVID-19. Int. J. Infect. Dis.

Holland, L.A., et al., 2020. An 81 nucleotide deletion in SARS-CoV-2 ORF7a identified from sentinel surveillance in Arizona (jan-Mar 2020). J. Virol.

Kumpornsin, K., Kochakarn, T., Chookajorn, T., 2019. The resistome and genomic reconnaissance in the age of malaria elimination. Dis. Model. Mech. 12 (12).

Minh, B.Q., et al., 2020. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. Mol. Biol. Evol. 37 (5), 1530–1534.

O'Toole, A.M., (2020) JT. Phylogenetic Assignment of Named Global Outbreak LINeages. [cited 2020 6 May]; Available from: https://github.com/hCoV-2019/pangolin.

Okada, P., et al., 2020. Early transmission patterns of coronavirus disease 2019 (COVID-19) in travellers from Wuhan to Thailand, January 2020. Euro Surveill. 25 (8).

Pongpirul, W.A., et al., 2020. Journey of a thai taxi driver and novel coronavirus. N. Engl. J. Med. 382 (11), 1067–1068.

Quick, J., 2020. nCoV-2019 Sequencing Protocol. Available from: https://www.protocols.io/view/ncov-2019-sequencing-protocol-bbmuik6w.

Quick, J., et al., 2017. Multiplex PCR method for MinION and Illumina sequencing of Zika and other virus genomes directly from clinical samples. Nat. Protoc. 12 (6), 1261–1276.

Rambaut, A., et al., 2020a. A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. Nat. Microbiol.

Rambaut, A., Holmes, E.C., Pybus, O.G., 2020b. A Dynamic Nomenclature for SARS-CoV-2 to Assist Genomic Epidemiology.

Simmonds, P., 2020. Rampant C–>U hypermutation in the genomes of SARS-CoV-2 and other coronaviruses: causes and consequences for their short- and long-term evolutionary trajectories. mSphere 5 (3).

Taylor, J.K., et al., 2015. Severe acute respiratory syndrome coronavirus ORF7a inhibits bone marrow stromal antigen 2 virion tethering through a novel mechanism of glycosylation interference. J. Virol. 89 (23), 11820–11833.