



## SHORT COMMUNICATION

# Phylogenomic analysis of SARS-CoV-2 from third wave clusters in Malaysia reveals dominant local lineage B.1.524 and persistent spike mutation A701V

Suppiah, J.<sup>1\*</sup>, Kamel, K.A.<sup>1</sup>, Mohd-Zawawi, Z.<sup>1</sup>, Afizan, M.A.<sup>1</sup>, Yahya, H.<sup>2</sup>, Md-Hanif, S.A.<sup>3</sup>, Thayan, R.<sup>1</sup>

<sup>1</sup>Virology Unit, Infectious Disease Research Centre, Institute for Medical Research, National Institutes of Health, Ministry of Health Malaysia, Selangor, Malaysia

<sup>2</sup>Disaster, Outbreak, Crisis and Emergency Management Sector, Surveillance Section, Disease Control Division, Ministry of Health Malaysia, Federal Government Administrative Centre, Putrajaya, Malaysia

<sup>3</sup>International Travel and Health Regulation Sector, Surveillance Section, Disease Control Division, Ministry of Health Malaysia, Federal Government Administrative Centre, Putrajaya, Malaysia

\*Corresponding author: jeyanthi@moh.gov.my

### ARTICLE HISTORY

Received: 31 May 2021  
Revised: 30 June 2021  
Accepted: 30 June 2021  
Published: 31 July 2021

### ABSTRACT

The emergence of a third wave of COVID-19 infection in Malaysia since September 2020 has led to imminent changes in public health prevention and control measures. As high as 96.2% of registered COVID-19 cases and 88.5% of confirmed deaths in Malaysia occurred during this third wave of infection. A phylogenomic study on 258 SARS-CoV-2 full genomes from February 2020-February 2021 has led to the discovery of a novel Malaysian lineage B.1.524. This lineage contains another spike mutation A701V that co-exists with the D614G spike mutation that was predominant in most of the third-wave clusters. The study provides vital genomic insights on the rapid spread of the SARS-CoV-2 variants in Malaysia in conjunction with the presence of a dominant SARS-CoV-2 lineage during the third wave of COVID-19 infection.

**Keyword:** COVID-19; phylogenomic; Malaysia; A701V; third wave.

### INTRODUCTION

Coronavirus disease 2019 (COVID-19) has been declared a global pandemic since March 2020 (European Centre for Disease Prevention and Control, 2020). Malaysia was among the earliest affected countries. The COVID-19 outbreaks in Malaysia are characterised by clusters, each named exclusively based on the location of the outbreak. The first wave of infection started on 24 January 2020 with the detection of 22 cases, of which 12 reported to have travelled to countries or regions with COVID-19 outbreaks (World Health Organization, 2020). Following that, the second wave started from 27 February 2020 until 19 September 2020. A total of 10,145 cases, 100 clusters, and 136 deaths were recorded during the second wave (Abdullah, 2020). The Sri Petaling Tabligh cluster that resulted from a mass religious gathering was a main contributor to the spike in COVID-19 cases during this phase (Malaymail, 2020). Some other major clusters also emerged towards the end of the second wave around August 2020 in the state of Kedah (Sivagangga, Tawar, Sala, Sungai, & Tembok clusters). Not long after the second wave settled, the third wave began when a major cluster was declared in Sabah, i.e. the Benteng Lahad Datu (Benteng LD) cluster in September 2020 (TheStar, 2020a). The cases from this cluster subsequently spread to Peninsular Malaysia, leading to the emergence of post-Benteng LD clusters.

Based on the data from the Crisis Preparedness Response Centre (CPRC) of the Ministry of Health Malaysia (MOH), a total of 290, 585 cases and 1000 deaths have been reported as contributed solely by the third wave of COVID-19 as of 28<sup>th</sup> February 2021 which is equivalent to 96.6.2% of all the COVID-19 cases in the country at that time. Many existing studies on epidemiological linking and mathematical modelling that forecast COVID-19 transmission (Gill *et al.*, 2020) have provided great assistance to the MOH in controlling the transmission of COVID-19 in Malaysia. However, it remains crucial to study the evolution of the circulating SARS-CoV-2 variants to monitor all the possible outcomes of the viral mutations, especially on the potential implications towards the community and population. To achieve this monitoring purpose, the genomic analysis must be performed in a timely manner. This study was conducted to analyse the SARS-CoV-2 full genomes obtained from the third wave clusters in Malaysia, with a particular focus on the Benteng- LD and post-Benteng-LD clusters in order to provide an insight on any possible evolution that has taken place.

### MATERIALS AND METHODS

The study received expedited ethical approval from the Medical Research and Ethics Committee (MREC) (NMRR-20-

884-54816). A total of 258 clinical specimens obtained from COVID-19 suspected patients from state hospitals and sent for routine diagnostic to our reference laboratory were included in this study. Viral nucleic acids were extracted from positive cases of COVID-19 using the QIAmp Mini Viral Nucleic Acid extraction kit (Qiagen, USA). The original specimens were selected based on a cycle threshold value of less than 25 (CT  $\leq$ 25) after routine detection assay of SARS-CoV-2 genome using the BGI Real-Time RT-PCR kit (Beijing, China). Specimens with CT >25 were subjected to viral isolation prior to selection for whole-genome sequencing. After that, the quantity of all the extracted viral RNA was assessed using Qubit RNA High Sensitivity assay (Invitrogen, USA) on Qubit 4 Fluorometer (Invitrogen, USA).

In the next step, the genomic library was prepared using the TruSeq Stranded Total RNA Library Prep Gold kit (Illumina, USA) before being sequenced with 2×75-bp paired-end library on a NextSeq500 platform (Illumina). Good quality sequencing reads that were obtained via BBDuk (BBTools version 38.57) trimming and filtering were assembled using MEGAHIT (version 1.2.8) to obtain contigs. Sequence similarity search for SARS-CoV-2-specific contig filtering was performed with BLASTN (version 2.9.0+). Then, the genome was annotated by aligning known SARS-CoV-2 protein sequences to complete genome sequences using Exonerate (Slater & Birney, 2005) whereas incomplete genomes were predicted using the Prokka pipeline (Seeman, 2014). The full genome sequences of SARS-CoV-2 were aligned with genomes from other countries available in the GISAID database using EMBOSS (Minh *et al.*, 2020). Lastly, a phylogenetic tree was constructed with MEGA-X using the Neighbour-Joining algorithm and bootstrap 1000x replication. All variants were annotated with mutation type using SnpEff (Cingolani *et al.*, 2012).

## RESULTS

A total of 258 SARS-CoV-2 full genomes encompassing COVID-19 cases from February 2020 to February 2021 were sequenced (GISAID accession numbers: EPI\_ISL\_430439-1424518). This included five SARS-CoV-2 genome sequences from the first wave, 53 from the second wave, and 200 from the third wave clusters respectively. Following the adaptation of the lineage classification described by Forster *et al.* (2020), the phylogenetic analysis of the Malaysian SARS-CoV-2 genomes revealed the presence of lineages B-Wuhan and B-derived (Figure 1). The clade L, lineage B-Wuhan made up the predominant circulating SARS-CoV-2 variants during the first wave and early second wave of COVID-19 outbreaks in Malaysia with an overall prevalence of 4.3% (n=11). From this point onwards, an evolution took place with the formation of clades and lineages. The circulating variants during the mid-second wave and third wave were mainly dominant lineages derived from B-Wuhan. Two major lineages that emerged during the second wave of infection included the clade O, lineage B.6.2 (4.3%, n=11) during the mid-second wave, and the clade GR, lineage B.1.1.354 (7.4%, n=19) at the end of the second wave. By this time, the ancestral lineage B-Wuhan was also found to have diminished. During the third wave of COVID-19, the dominance of a single clade GH, lineage B.1.524 was observed, with a high prevalence of 58.5% (n=151).

Additionally, two dominant spike mutations, i.e. D614G and A701V were constantly found among the majority of the SARS-CoV-2 circulating in Malaysia with an overall prevalence of 86.4% (223/258) and 62.0% (n=160/258) respectively. Apart from D614G, the emergence of the A701V mutation in the

spike gene in lineage B.1.524 was predominant during the third wave clusters. The prevalence of this mutation was detected in 62.0% of all the sequenced strains in this study, of which 80.0% were from the third wave clusters. Unlike D614G, the A701V variant was not observed in any genomes of SARS-CoV-2 prior to the Benteng-LD cluster (Figure 2).

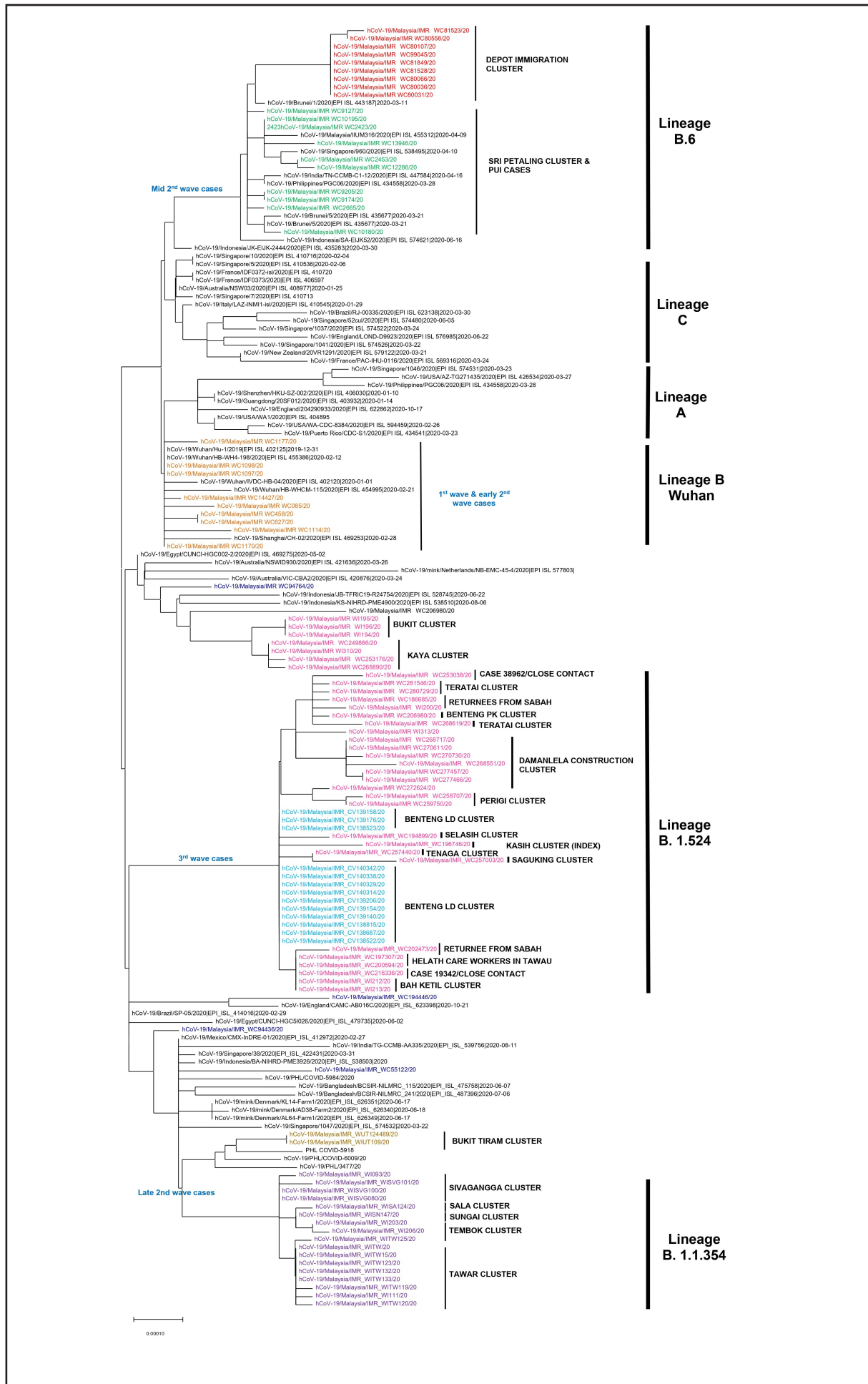
## DISCUSSION

The study highlights the emergence of a local lineage, namely B.1.524, and the conferred spike mutation A701V of SARS-CoV-2 in Malaysia. The Pango lineage nomenclature had officially classified the B.1.524 as a Malaysian lineage (Pango lineages, 2021). The lineage was first discovered among cases from the Benteng-LD cluster in Sabah in early September 2020. The Benteng-LD cluster was identified following the screening of new detainees locked up at the Lahad Datu District Police Headquarters (IPD), Sabah on 1<sup>st</sup> September 2020. Within ten days, up to 500 positive cases were detected among prisoners and the general public from this cluster, with the majority being Malaysians (Malaysiakini, 2020). A total of 22 SARS-CoV-2 full genome sequences from samples collected from this cluster on 1<sup>st</sup> September 2020 are available in GISAID (EPI\_ISL\_718298 – 718304). After the initial discovery in Sabah, it spread rapidly and by the end of September, the B.1.524 lineage was observed among the clusters in Peninsular Malaysia, mainly among returnees from Sabah after the state general election. The first case infected by lineage B.1.524 in Peninsular Malaysia was traced to a sample collected on 28 September 2020 (EPI\_ISL\_728157).

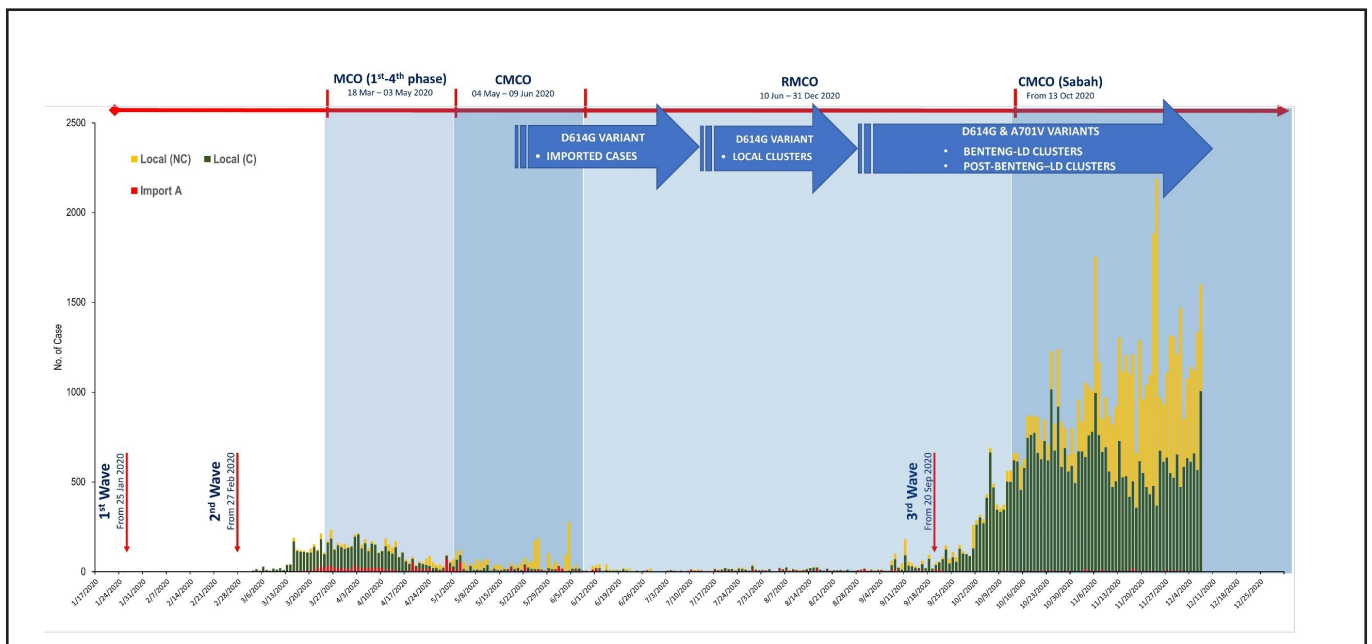
By focusing on the 151 full genome sequences of lineage B.1.524 from the third wave of infection in the phylogenetic tree (Figure 2), evolutionary proximity can be observed between the SARS-CoV-2 variants from the Benteng LD and post-Benteng LD clusters in Malaysia. Thus, it is highly suggestive that the Benteng LD variant could be the parental strain responsible for the emergence of the third wave clusters. The emergence of spike mutation A701V in lineage B.1.524 among the vast majority of third wave cases was significantly concordant with the sudden spike of COVID-19 cases. Furthermore, our postulation was supported by the data from GISAID, in which the lineage B.1.524 that harboured the mutations A701V and D614G was first identified in large numbers in Malaysia during early September 2020 before being deposited in GISAID in November 2020, thus demonstrating a high possibility that the lineage originated from Malaysia. Even though the lineage was reported in other countries, it was only in a much later period. As of February 2021, the deposition of lineage B.1.524 in GISAID originated from other countries, including Singapore (n=60), Australia (n=2), and Cambodia (n=1).

The persistent co-circulation of the A701V with D614G highly indicates that these two spike mutations have enhanced the survival fitness of the Malaysian lineage B.1.524. Recent studies demonstrated that the D614G transmitted significantly faster and produced higher infectious titres (Hou *et al.*, 2020; Plante *et al.*, 2020). Based on our data, the spike mutation D614G was first observed in the SARS-CoV-2 genomes isolated from imported cases. After that, they were persistently transmitted to second and third wave clusters including Sivagangga, Tawar, Sungai, Tembok, Benteng LD, and post-Benteng LD.

The impact of the lineage containing the D614G and A701V spike mutations in Malaysia was evident through the enormous transmission of COVID-19 within a short period. From the very early stage of the second wave clusters, our genome sequencing findings have cautioned the existence



**Figure 1.** Phylogenetic tree showing the dominance of B.1.524 lineage in Malaysia (representative sequences) and the evolutionary relationship among various clusters in Malaysia represented by different font colours.



**Figure 2.** The overview of COVID-19 cases and transition of important spike mutations during the three major waves of infection in Malaysia (MCO: Movement Control Order; CMCO: Conditional Movement Control Order; RMCO: Recovery Movement Control Order; NC: Non-citizen; C: Citizen).

of the D614G in a timely manner, hence leading to prompt decision making on contact tracing and containment efforts (TheStar, 2020b). However, the Benteng LD cluster failed to be contained and led to the subsequent outbreaks of third wave clusters. Some mutations were efficiently preserved in subsequent clusters and found to be capable of providing enhanced fitness to the virus, thus indicated that positive selection might have taken place. A previous study has reported that positively selected mutants can produce the strongest transmission in a dense population with a large number of immunologically naive individuals (Troy *et al.*, 2020). Another study finding revealed that the A701V spike mutation is located at the furin cleavage site which is also found in the highly transmissible Beta variant (B.1.351) (Annavaajhala *et al.*, 2021). The furin cleavage site is an important determinant of SARS-CoV-2 transmission whereby mutations emerging at this site potentially increase the cleavage of S1/S2 subunits of the spike protein and enhances the cell entry (Peacock *et al.*, 2021). Thus, from the genomic aspect, a possibility can be linked to the role of the novel A701V mutation whereby it may have accelerated the transmission of the Benteng LD and post-Benteng LD clusters, causing the virus to become fitter than the D614G variant.

Several strengths are inherent in this study. To the best of the researcher's knowledge, this is the first study that elucidates the presence of the dominant lineage and spike mutation during the third wave of the COVID-19 pandemic in Malaysia. Thus, it provides an added value to the existing study that reported on the dominance of lineage B.6 during the early phase of the pandemic (Chong *et al.*, 2020). Additionally, a large sample size that originated solely from our laboratory in the Institute for Medical Research (IMR) was applied in the study. As the national reference laboratory for COVID-19 genomic surveillance, IMR is the largest contributor to the genomic surveillance of SARS-

CoV-2 in Malaysia. Thus, this study analysed the samples from multiple sentinels in all the states of Malaysia. Therefore, the study findings are meaningful and nationally representative of the current SARS-CoV-2 circulation in Malaysia.

## CONCLUSION

This study focused on identifying the dominant lineage and spike mutations that circulated during the emergence of the third wave COVID-19 infection in Malaysia. The emergence of A701V spike mutation in lineage B.1.524 among the majority of the third wave clusters in Malaysia is an impactful observation. This analysis provided vital insight into the evolutionary impact of SARS-CoV-2 mutations on the enormous spike in COVID-19 cases in Malaysia during this period. Even though this mutation was also observed in a few SARS-CoV-2 genomes from other countries, its impact on the transmission was not highlighted. The findings from this study emphasises the need for continuous monitoring of the A701V spike mutation.

## ACKNOWLEDGEMENT

The authors would like to thank the Director General of Health Malaysia for granting permission to publish this paper. This work was funded by the Ministry of Health Malaysia Research Grant (NMRR-20-884-54816; 20-039). We also acknowledge the contribution of Malaysian Genome Institute for the submission of four genome sequences to GISAID (EPI\_ISL\_455790–455793) from this study.

## Conflict of interest

The authors declare no conflict of interest.



## REFERENCES

- Abdullah, N.H. (2020). Situasi terkini jangkitan 2019 novel Coronavirus (2019-nCoV) di Malaysia [Press release]. <https://kpkesihatan.com/2020/09/19/kenyataan-akhbar-kpk-19-september-2020-situasi-semasa-jangkitan-penyakit-coronavirus-2019-covid-19-di-malaysia/>. Accessed 30 September 2020.
- Annavajhala, M.K., Mohri, H., Zucker, J.E., Sheng, Z., Wang, P., Gomez-Simmonds, A., Ho, D.D. & Uhlemann, A.C. (2021). A novel SARS-CoV-2 variant of concern, B.1.526, identified in New York. *medRxiv*. <https://doi.org/10.1101/2021.02.23.21252259>
- Chong, Y.M., Sam, I.C., Chong, J., Kahar Bador, M., Ponnampalavanar, S., Syed Omar, S.F., Kamarulzaman, A., Munusamy, V., Wong, C.K. & Jamaluddin, F.H. (2020). SARS-CoV-2 lineage B.6 was the major contributor to early pandemic transmission in Malaysia. *PLoS Neglected Tropical Disease* **14**: e0008744. <https://doi.org/10.1371/journal.pntd.0008744>
- Cingolani, P., Platts, A., Wang, L.L., Coon, M., Nguyen, T., Wang, L., Land, S.J., Lu, X. & Ruden, D.M. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly* **6**: 80-92. <https://doi.org/10.4161/fly.19695>
- European Centre for Disease Prevention and Control. (2020). COVID-19 situation updates worldwide. <https://www.ecdc.europa.eu/en/geographical-distribution-2019-ncov-cases>. Accessed 22 Nov 2020.
- Forster, P., Forster, L., Renfrew, C. & Forster, M. (2020). Phylogenetic network analysis of SARS-CoV-2 genomes. *Proceedings of the National Academy of Sciences* **117**: 9241-9243. <https://doi.org/10.1073/pnas.2004999117>
- Gill, B.S., Jayaraj, V.J., Singh, S., Mohd Ghazali, S., Cheong, Y.L., Md Iderus, N.H., Sundram, B.M., Aris, T., Mohd Ibrahim, H. & Hong, B.H. (2020). Modelling the effectiveness of epidemic control measures in preventing the transmission of COVID-19 in Malaysia. *International Journal of Environmental Research and Public Health* **17**: 5509. <https://doi.org/10.3390/ijerph17155509>
- Hou, Y.J., Chiba, S., Halfmann, P., Ehre, C., Kuroda, M., Dinnon, K.H., Leist, S.R., Schäfer, A., Nakajima, N. & Takahashi, K. (2020). SARS-CoV-2 D614G variant exhibits efficient replication ex vivo and transmission in vivo. *Science* **370**: 1464-1468. <https://doi.org/10.1126/science.abe8499>
- Malaymail. (2020). Malaysia's biggest COVID-19 cluster linked to Sri Petaling tabligh event ends today after four months. Malaymail. <https://www.malaymail.com/news/malaysia/2020/07/08/malaysias-biggest-covid-19-cluster-linked-to-sri-petaling-tabligh-event-end/1882691>. Accessed 30 September 2020.
- Malaysiakini. (2020). Benteng cluster swells as 182 new Covid-19 cases reported. Malaysiakini. <https://www.malaysiakini.com/news/542247>. Accessed 20 January 2021.
- Minh, B.Q., Schmidt, H.A., Chernomor, O., Schrempf, D., Woodhams, M.D., von Haeseler, A. & Lanfear, R. (2020). IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. *Molecular Biology and Evolution* **37**: 1530-1534. <https://doi.org/10.1093/molbev/msaa015>
- Pango Lineages. (2021). Available at: [https://cov-lineages.org/lineages/lineage\\_B.1.524.html](https://cov-lineages.org/lineages/lineage_B.1.524.html). Accessed 19 February 2021.
- Peacock, T.P., Goldhill, D.H., Zhou, J., Baillon, L., Frise, R., Swann, O.C., Kugathasan, R., Penn, R., Brown, J.C. & Sanchez-David, R.Y. (2021). The furin cleavage site in the SARS-CoV-2 spike protein is required for transmission in ferrets. *Nature Microbiology* **6**: 899-909. <https://doi.org/10.1038/s41564-021-00908-w>
- Plante, J.A., Liu, Y., Liu, J., Xia, H., Johnson, B.A., Lokugamage, K.G., Zhang, X., Muruato, A.E., Zou, J. & Fontes-Garfias, C.R. (2021). Spike mutation D614G alters SARS-CoV-2 fitness. *Nature* **592**: 116-121. <https://doi.org/10.1038/s41586-020-2895-3>
- Seeman, T. (2014). Prokka: rapid prokaryotic genome annotation. *Bioinformatics* **30**: 2068-2069. <https://doi.org/10.1093/bioinformatics/btu153>
- Slater, G.S.C. & Birney, E. (2005). Automated generation of heuristics for biological sequence comparison. *BMC Bioinformatics* **6**. <https://doi.org/10.1186/1471-2105-6-31>
- TheStar. (2020a). Benteng Lahad Datu cluster spikes again with 62 new cases. TheStar. <https://www.thestar.com.my/news/nation/2020/09/08/benteng-lahad-datu-cluster-spikes-again-with-62-new-cases>. Accessed 10 November 2020.
- TheStar. (2020b). Dr Noor Hisham: Tawar, Sivagangga clusters likely linked as both carry D614G mutation. TheStar. <https://www.thestar.com.my/news/nation/2020/08/27/dr-noor-hisham-tawar-sivagangga-clusters-likely-linked-as-both-carry-d614g-mutation>. Accessed September 2020.
- Troy, D., Sylvain, G., Sébastien, L. & Sarah, P.O. (2020). On the evolutionary epidemiology of SARS-CoV-2. *Current Biology* **30**: R849-R857. <https://doi.org/10.1016/j.cub.2020.06.031>
- World Health Organization. (2020). COVID-19 Situation Overview in Malaysia. <https://www.who.int/malaysia/internal-publications-detail/covid-19-in-malaysia-situation-report-01>. Accessed 20 July 2020.