INNSPUB

International Journal of Biosciences | IJB |
ISSN: 2220-6655 (Print), 2222-5234 (Online)
http://www.innspub.net
Vol. 20, No. 5, p. 95-102, 2022

RESEARCH PAPER OPEN ACCESS

Genetic diversity of SARS-CoV-2 Variants' spike gene after 4th wave in Cambodia

Quan Ke Thai[1], Phuoc Huynh[4], Vu Nguyen Quoc[2], Yen Le Thi[3], Huyen Nguyen Thi Thuong[5*]

[1]*Saigon University, 273 An Duong Vuong, Ward 3 District 5, Ho Chi Minh city, Vietnam*

[2]*DSI1181, Saigon University, 273 An Duong Vuong, Ward 3 District 5, Ho Chi Minh city, Vietnam*

[3]*DSI1191, Saigon University, 273 An Duong Vuong, Ward 3 District 5, Ho Chi Minh city, Vietnam*

[4]*VNU HCMC University of Science, 227 Nguyen Van Cu, Ward 4 District 5, Ho Chi Minh city, Vietnam*

[5]*Department of Biology, HCMC University of Education, 280 An Duong Vuong Ward 4 District 5, Ho Chi Minh city, Vietnam*

Key words: Cambodia, Covid-19, Genetic diversity, SARS-CoV-2, Spike gene.

http://dx.doi.org/10.12692/ijb/20.5.95-102 Article published on May 24, 2022

Abstract

SARS-CoV-2 caused millions of deaths and hundreds of millions of infections over the world. While many countries and regions were affected strongly by this new virus, Cambodia showed success in limiting the death and infection before going bad in the fourth wave of the pandemic. In order to understand this success, this study examines the genetic diversity of the Spike gene of SARS-CoV-2 variants in Cambodia. The results indicated the high-frequency haplotype and nucleotide diversity in Cambodia, respectively 0.8669 ± 0.0090 and 0.002761 ± 0.001392. In the S gene, 278 nucleotide polymorphisms were recorded, in which mutation A23403G (D614G) accounted for the highest frequency, corresponding to 97.50%. The gene regions coding for the corresponding positions in the protein are NTD, RBD, and Furin S1/S2 cleavage site (FCS), which are the regions where mutations occur. The genetic diversity of Spike gene of SARS-CoV-2 variants really supported the observation of the success in disease control.

* **Corresponding Author:** Huyen Nguyen Thi Thuong ✉ huyenntth@hcmue.edu.vn

Introduction

A recently emerging virus, SARS-COV-2, which has caused millions of deaths and hundreds of millions of infections, is a great danger and challenge to global health, especially for countries with low-income, underdeveloped healthcare. SARS-COV-2 belongs to the Betacoronavirus genus (β-CoV or Beta-CoV), is one of four genera of coronaviruses of the family Coronaviridae, contained in the nucleocapsid intestine is a single-stranded +RNA about 30kb (Gupta, Charron *et al.*, 2020) that encodes for four structural proteins. The main structures are Membrane (M), Envelope (E), Spike (S), and Nucleocapsid (N). The S gene encoding for S protein plays a crucial role in viral processes such as infection, antigen presentation, and interaction with neutralizing antibodies (Ou, Liu *et al.*, 2020) and is an important target in vaccine strategy.

According to the WHO Working definition, the VOC variants possess different mutations in the S gene region as the phenotypic characteristics of the virus. Mutations in the S gene increase transmissibility, virulence, or change in clinical disease presentation and can decrease the effectiveness of diagnostics tests, vaccines, or therapeutics through glycosylation modification (Guruprasad, 2021). Therefore, tracking different mutations in the S gene is necessary for the context that countries worldwide need to quickly recover their economies in the "new normal" period after COVID-19.

Despite the substantial spread of COVID-19 in countries worldwide, such as China, the United States, and Europe, Cambodia has initially succeeded in controlling the domestic infection in the early stages of the pandemic even though this is a country with an underdeveloped economy and health. Cambodia's anti-epidemic effectiveness is due to the government's quick response, quickly imposing blockade orders, closing borders, restricting movement into the country as negative COVID-19, undergoing PCR testing on arrival, stay in quarantine for 14 days (Nov, Hyodo *et al.*, 2021). These include travel restrictions for people from Thailand and Laos, and domestic and international flights to Malaysia, Indonesia, and the Philippines were suspended (Chiti, Stefani *et al.*, 2003). In addition, the Cambodian Ministry of Health conducts rapid training of highly qualified staff (called Rapid Response Teams (RRT)) to serve the lines with low medical backgrounds, spanning all 25 provinces of Cambodia (Nit, Samy *et al.*, 2021). Another critical factor in the effectiveness of Cambodia's disease control is that the disease control situation in neighboring countries Laos and Vietnam has achieved great success (Nit, Samy *et al.*, 2021), especially in Vietnam with 99 days without any recorded cases of community infection (Thai, Rabaa *et al.*, 2021), as well as medical support from WHO and other countries (Nit, Samy *et al.*, 2021) (Pande, 2004). Cambodia encountered two outbreaks, with the first recorded case on January 27, 2020, and until February 15, 2021; the Kingdom recorded only 479 positive cases of SARS-CoV-2, of which up to 83% are imported cases (Nov, Hyodo *et al.*, 2021). Despite efforts from the government, shortages of food and expertise, medical staff, and the introduction of new variants with solid infectious properties.

The consequence was subsequent outbreaks that made disease control worse in Cambodia as hospitals became overwhelmed, food supplies were inconsistent, and 83% faced food shortages (Tatum, 2021); during the same period, the influenza A(H_3N_2) outbreak also broke out in this country (Siegers Jurre, Dhanasekaran *et al.*) leading to a public health crisis and epidemiological control of the patient. The retardation in diagnostic testing and genome sequencing of SARS-CoV-2 is also one of the main reasons for the intense outbreak when the novel variants appeared in Cambodia with strong infectious potential stronger than previous ancestors.

The principal explanation is due to the lack of conditions for people to carry out diagnostic tests because the quarantine orders limit their income (Tatum, 2021), the sparse number of medical professionals (Kosaka, Kobashi *et al.*, 2021), the telemedicine system is not prevalent, the qualifications of medical staff in remote areas are still

limited, and need to be improved (Garcia-Beltran, Lam *et al.*, 2021). Another reason is the lack of medical supplies and highly qualified specialists in sequencing and bioinformatics (Manning, Bohl *et al.*, 2020). PCR testing is performed only at two specialist institutes in Phnom Penh (Iwamoto, Tung *et al.*, 2020), further stalling case identification. The consequence is delayed recording of case information for case-contact tracing and tracking viral genome to identify the source of infection. In this study, the genetic diversity of SARS-CoV-2 spike genes reported in Cambodia were examined to reveal the genetic diversity of SARS-CoV-2, supporting the success of the medical system in the prevention of the SARS-CoV-2 spreading in Cambodia.

Materials and methods

Data collection

The sequences of SARS-CoV-2 were extracted from the GISAID. The extracted information includes accession number, place of sample collection, gender, collection date, and classification of GISAID, PANGO Lineages, Total of 1120 whole genome of viruses were submitted on GISAID from January 27, 2020, to September 18, 2021.

Multiple sequence alignment (MSA) analysis

Separation of the S gene region from the whole genome based on the well-defined S gene region on the reference sequence (accession number: NC_045512.2). MEGA software (Tamura, Stecher *et al.*, 2021) performed MAS to identify nucleotide mutations and protein substitutions encoded by the corresponding codon.

Genetic diversity analysis

Genetic diversity parameters, including nucleotide and haplotype diversity, were calculated using Arlequin software ver 3.5.2.2 (Excoffier and Lischer, 2010).

Haplotype network

POPART software ver 1.7 (Leigh and Bryant, 2015) was used to build the haplotype network to analyze SNVs occurring in the S gene.

Results

Genetic diversity

The genetic diversity analysis detected 279 haplotypes of the S gene in Cambodia. The haplotypes of the Alpha variants had the most considerable frequency, followed by Delta variants, B.1 group, and Rwanda, respectively 67.50%, 26.34%, 3.66%, 1.25% (Fig. 1).

Examination of the S gene indicates haplotype diversity of 0.8669 ± 0.0090 with an average nucleotide distance of 10.551664 ± 4.811940, demonstrating that multiple variants have immigrated to this country.

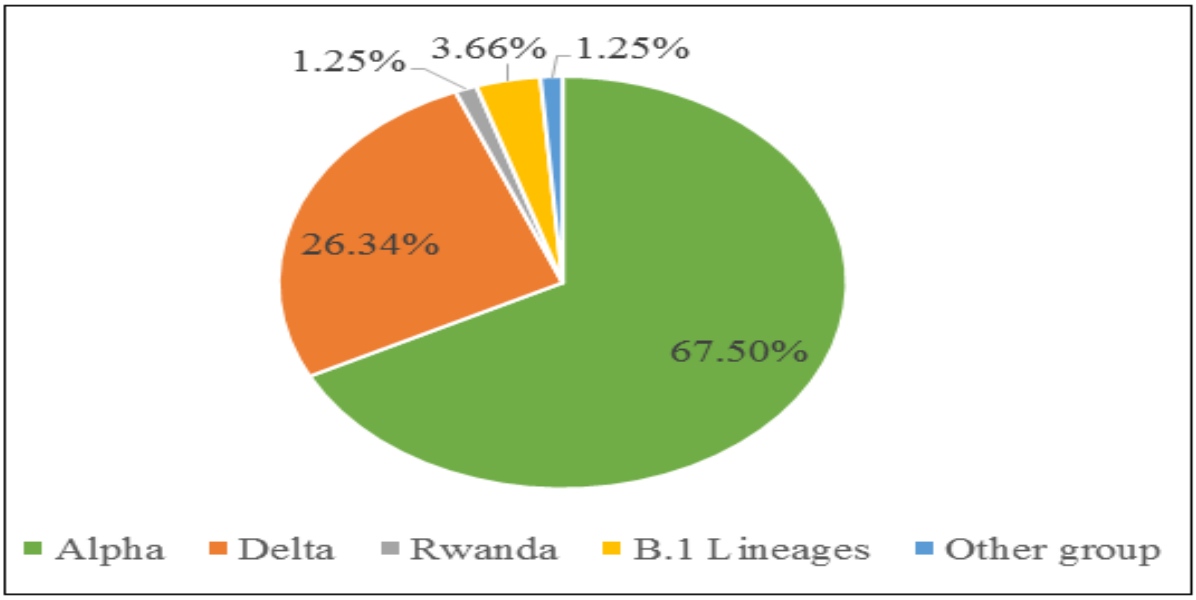

Fig. 1. SARS-CoV-2 variants composition in Cambodia.

The analysis of compositions indicates the length of the S gene is 3822 nucleotides encoding for a protein sequence of 1273 amino acids length, and nucleotide composition has the characteristic that G+C is approximately 37.29%. The T nucleotides were the most dominant, followed by A, C, and G, respectively 33.28%, 29.44%, 18.85%, and 18.44%. The results of nucleotide diversity recorded multiple mutations in the S gene with the nucleotide diversity of 0.002761 ± 0.001392. The identification of 278 polymorphic sites nucleotide. In which 261 substitutions, including 167 sites with transitions, 99 transversions, and 18 deletions. The missense mutations have a rate of 58.28%, and C > T transitions have the highest occurrence rate at 34.39% (Fig. 2). The isolated viruses in Cambodia contain A23403G mutation in the S gene is considered the most predominant, with a frequency of 97.50%.

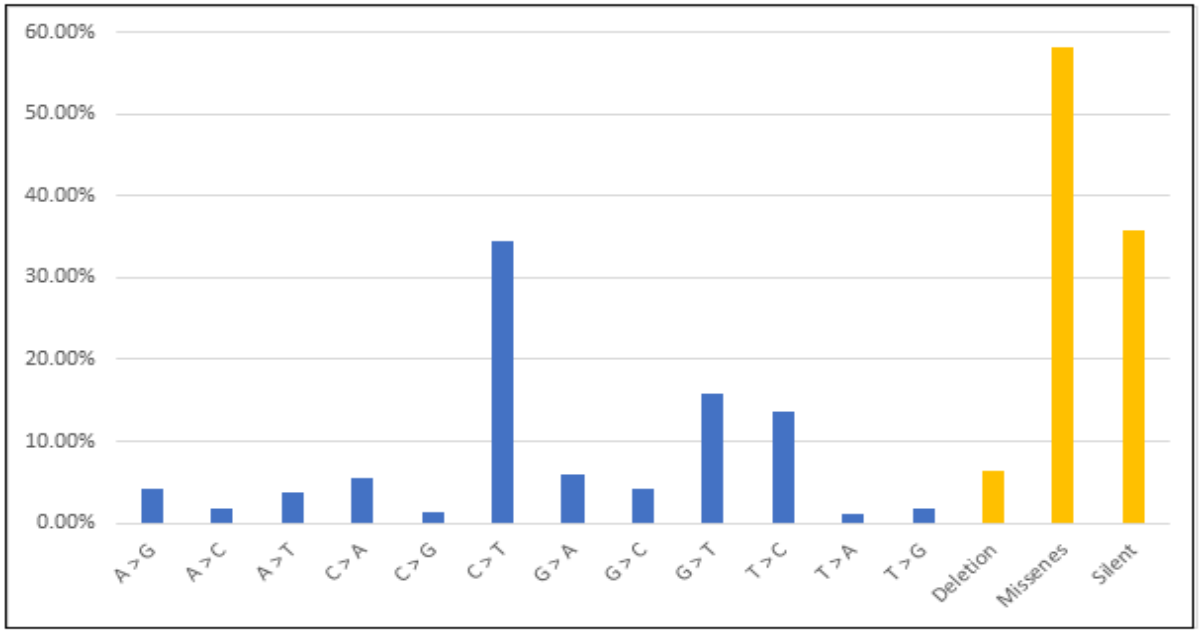

Fig. 2. Major mutations in Spike genes.

The distribution position of mutations in the domains of the S gene indicates that the mutations appeared scattered across the gene. The mutation frequency distribution of S1 and S2 subunits are relatively equal, respectively, 53.33% and 46.67%. The mutations mainly appeared in the S1 region, in which the NTD region carried the most deletion. The region coding for RBD has the high-frequency mutation, notably A23063T (N501Y), G23012A (E484K), C22995A (T478K), G22927T (L455F), and T22917G (L452R) (Fig 3). The mutations with the highest frequency and qualified for changing the infectivity of the virus appeared around the FCS region (Fig. 3).

Haplotype network

The haplotype network of Cambodia exhibits several significant central haplotypes (Fig. 4), and the interconnected network nodes radiate a star pattern; the haplotypes differ only by 1 to 2 nucleotides from the central node. Especially for large clusters, the haplotypes form new branches and differ only 1 to 2 nucleotides from the haplotype in the cluster. The results show that the intrinsic haplotypes of these clusters are closely related and have intense variation in the S gene. In addition, there is the import of other haplotypes into Cambodia when there is a significant nucleotide difference. The haplotypes belonging to G614Group have a substantial nucleotide difference of 2 to 7 nucleotides.

Discussion

We reported a high prevalence of missenses mutations in the S gene of SARS-CoV-2 in Cambodia and the comparable prevalence of these mutations in the regions coding two subunits. However, mutations with a high frequency of occurrence are distributed

unevenly. Critical mutations in the S gene mainly appear in the region coding for the S1 subunit and FCS. The RBD region in the S gene records high-frequency mutation. In addition, detecting a combination of three mutations in RBD [L452R, T478K, N501Y] is consistent with the prediction in the study of Jiahui Chen (Chen, Wang *et al.*, 2021). At the S2 subunit, the region around the FCS also appeared with mutations that can help the virus facilitate infection and replication. In this study, multiple mutations with the highest frequency and improved the transmission of virals appeared near the FCS, including C23271A (A570D), A23403G (D614G), C23604A/G (P681H/R), C23709T (T716I), respectively 67.32%, 97.50%, 67.86%, 28.04%, 67.68%, which D614G substitution is the most common mutation and is also a fundamental mutation in the evolution of SARS-CoV-2.

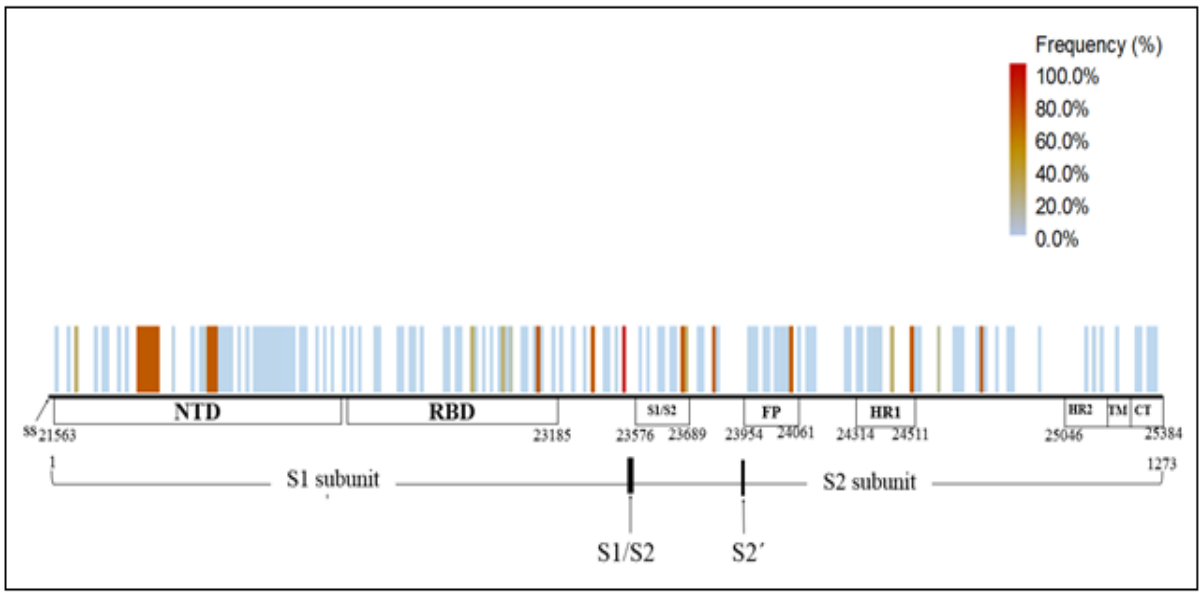

Fig. 3. Polymorphism positions over the S gene.
(*Each column represents each mutation; SS: Signal sequences; NTD: N-terminal domain; RBD: Receptor binding domain; S1/S2: Furin S1/S2 cleavage site; FP: Fusion peptide; HR1: heptad repeat 1; HR2: heptad repeat 2; TM: transmembrane domain; CT: cytoplasmic tail*).

In Cambodia, since it first appeared on March 11, 2020, this mutation has become dominant in the population during the following period. Examinations on the D614G substitution indicate the variants maintaining this mutation ability increased the viral's infectivity (Yurkovetskiy, Wang *et al.*, 2020) (Korber, Fischer *et al.*, 2020). Notably, P681H and P681R are two mutations that have been shown to favor the virus during infection, in which P681R enhances the virulence of the Delta and is highly conserved in this variant are also present in Cambodia. In addition, the isolated virus in Cambodia also recorded G23401T (Q613H) substitution as the concentration when the function of this substitution is comparable to that of D614G (Lubinski, Frazier *et al.*, 2021). Thereby the FCS region plays an important role in the infection ability of SARS-CoV-2. Mutations appearing in this region will facilitate virus infection into host cells through enhanced spike protein cleavage.

The results of the genetic diversity analysis can demonstrate that, at first, the variants in Cambodia were closely related, and later, the entering of multiple virals into the country. The pieces of information support this statement when the actuality that until February 15, 2021, Cambodia recorded 479 cases of SARS-CoV-2 infection, of which 396 (83%) were imported cases. The stabilization of the pandemic situation in Vietnam and Laos can also highlight Cambodia's first two waves of epidemics. In addition, the haplotype network analysis shows relative variation in the S gene. We found that the

haplotypes in the same group of Cambodia are closely related (from 1 to 2 nucleotides). However, despite the government's measures, the third wave of outbreaks has worsened the epidemic situation when the highest daily infection cases recorded was 1130 cases. The introduced variants with high transmissibility were likely to be the leading cause of the outbreak.

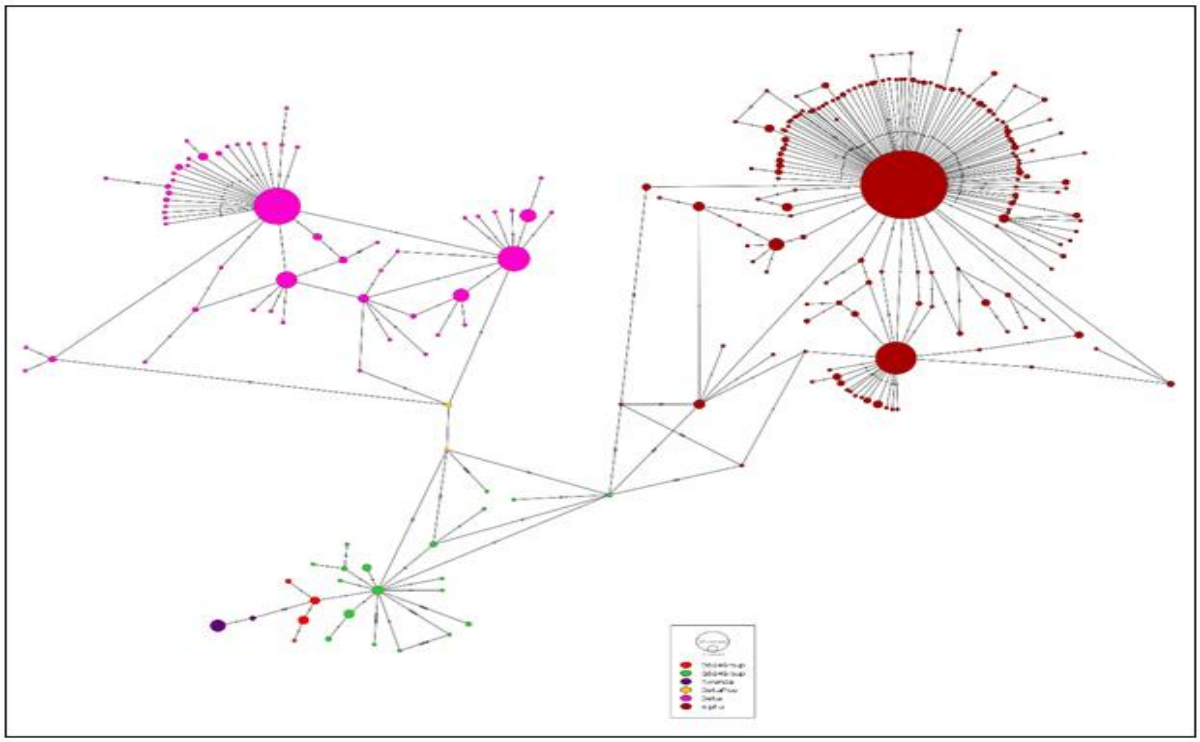

Fig. 4. Minimum-spanning network of spike genes.

Consequently, there is intense community transmission of these groups when numerically dominant in the population. However, the middle nodes in the network considerably differ from the closest nodes (from 2 to 7 nucleotides). Thereby, it is probable to identify these intermediate haplotypes in the future. In the haplotype network, we also recognized the importance of the A230403G (D614G) substitution when this is between nodes – an important bridge in the network.

Conclusions

In conclusion, this study provided information on the haplotype and nucleotide diversity of the S gene and revealed amino acid substitutions in the spike protein of SARS-CoV-2 in Cambodia. Most of Cambodia's high-frequency mutations in the S gene region are present in the RBD and FCS regions. The genetic diversity parameters of the S gene in Cambodia were expressed at a high level and had a sizeable mean nucleotide difference. The A23403G (D614G) substitution dominates both in number and importance in the evolution of SARS- CoV-2. The results suggest rapid genetic evolution in the S gene of SARS-CoV-2 through the accumulation of mutations during intense community transmission. Therefore, it is necessary to continuously monitor the genetic variation of the S gene in the future.

Reference

Chen J, Wang R. 2021. Review of the mechanisms of SARS-CoV-2 evolution and transmission. ArXiv: arXiv:2109.08148v08141.

Chiti F, Stefani M. 2003. Rationalization of the effects of mutations on peptide and protein aggregation rates. Nature **424(6950),** 805-808.

Excoffier L, Lischer HE. 2010. Arlequin suite ver 3.5: a new series of programs to perform population

genetics analyses under Linux and Windows. Molecular Ecology Resources **10(3),** 564-567. http://dx.doi.org/10.1111/j.1755-0998.2010.02847.x.

Garcia-Beltran WF, Lam EC. 2021. Multiple SARS-CoV-2 variants escape neutralization by vaccine-induced humoral immunity. Cell **184(9),** 2372-2383.e2379. http://dx.doi.org/10.1016/j.cell.2021.03.013.

Gupta R, Charron J. 2020. SARS-CoV-2 (COVID-19) structural and evolutionary dynamicome: Insights into functional evolution and human genomics. Journal of Biological Chemistry **295(33),** 11742-11753. http://dx.doi.org/10.1074/jbc.RA120.014873.

Guruprasad L. 2021. Human SARS CoV-2 spike protein mutations. Proteins **89(5),** 569-576. http://dx.doi.org/10.1002/prot.26042.

Iwamoto A, Tung R. 2020. Challenges to neonatal care in Cambodia amid the COVID-19 pandemic. Global Health & Medicine **2(2),** 142-144. http://dx.doi.org/10.35772/ghm.2020.01030.

Korber B, Fischer WM. 2020. Tracking Changes in SARS-CoV-2 Spike: Evidence that D614G Increases Infectivity of the COVID-19 Virus. Cell **182(4),** 812-827 e819. http://dx.doi.org/10.1016/j.cell.2020.06.043.

Kosaka M, Kobashi Y. 2021. Lessons from COVID-19's impact on medical tourism in Cambodia. Public Health in Practice **2,** 100182. http://dx.doi.org/10.1016/j.puhip.2021.100182.

Leigh JW, Bryant D. 2015. popart: full-feature software for haplotype network construction. Methods in Ecology and Evolution **6(9),** 1110-1116. https://doi.org/10.1111/2041-210X.12410.

Lubinski B, Frazier LE. 2021. Spike protein cleavage-activation mediated by the SARS-CoV-2 P681R mutation: a case-study from its first appearance in variant of interest (VOI) A.23.1 identified in Uganda. bioRxiv : the preprint server for biology: 2021.2006.2030.450632. http://dx.doi.org/10.1101/2021.06.30.450632.

Manning JE, Bohl JA. 2020. Rapid metagenomic characterization of a case of imported COVID-19 in Cambodia. bioRxiv : the preprint server for biology: 2020.2003.2002.968818. http://dx.doi.org/10.1101/2020.03.02.968818.

Nit B, Samy AL. 2021. Understanding the Slow COVID-19 Trajectory of Cambodia. Public Health in Practice **2,** 100073. http://dx.doi.org/10.1016/j.puhip.2020.100073.

Nov T, Hyodo T. 2021. Impact of the third wave of the COVID-19 pandemic and interventions to contain the virus on society and patients with kidney disease in Cambodia. Renal Replacement Therapy **7(1),** 53. http://dx.doi.org/10.1186/s41100-021-00372-6.

Ou X, Liu Y. 2020. Characterization of spike glycoprotein of SARS-CoV-2 on virus entry and its immune cross-reactivity with SARS-CoV. Nature Communications **11(1),** 1620. http://dx.doi.org/10.1038/s41467-020-15562-9.

Pande VS. 2004. A universal TANGO? Nature Biotechnology **22(10),** 1240-1241.

Siegers Jurre Y, Dhanasekaran V, Genetic and Antigenic Characterization of an Influenza A(H3N2) Outbreak in Cambodia and the Greater Mekong Subregion during the COVID-19 Pandemic. 2020. Journal of Virology **95(24),** e01267-01221. http://dx.doi.org/10.1128/JVI.01267-21.

Tamura K, Stecher G. 2021. MEGA11: Molecular Evolutionary Genetics Analysis Version 11. Molecular Biology and Evolution **38(7),** 3022-3027. http://dx.doi.org/10.1093/molbev/msab120.

Tatum M. 2021. Cambodia ends controversial

COVID-19 restrictions. Lancet (London, England) 397(10289), 2035-2035. http://dx.doi.org/10.1016/S0140-6736(21)01196-X.

Thai PQ, Rabaa MA. 2021. The First 100 Days of Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) Control in Vietnam. Clinical Infectious Diseases **72(9),** e334-e342. http://dx.doi.org/10.1093/cid/ciaa1130.

Yurkovetskiy L, Wang X. 2020. Structural and Functional Analysis of the D614G SARS-CoV-2 Spike Protein Variant. Cell **183(3),** 739-751 e738. http://dx.doi.org/10.1016/j.cell.2020.09.032