DETECTION OF SINGLE NUCLEOTIDE POLYMORPHISMS (SNPs) IN A SARS-CoV-2 VIRUS STRAIN IN VIETNAM

Tran Xuan Thach[1], Bui Thi Thuy Duong[1], Le Van Truong[1], Nguyen Thi Hoa[1], Pham Thi Lanh[1], Nguyen Thai Uy[1], Le Quynh Mai[2], Dang Duc Anh[2], Dinh Duy Khang[1], Dong Van Quyen[1,*]

[1]Institute of Biotechnology, VAST, Vietnam
[2]The National Institute of Hygiene and Epidemiology, Ha Noi, Vietnam

ABSTRACT

The severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is the causative agent of a pandemic of acute respiratory disease in humans. This pandemic has now spread worldwide and caused more than 2,000,000 deaths by 15 January 2021. The complexity and the ongoing pandemic of coronavirus SARS-CoV-2 make it difficult to control the disease. Sequence analysis on some open reading frames (ORFs) of a SARS-CoV-2 virus strain isolated in Vietnam in February 2020 (SARS-CoV-2/NIHE/human/2020/VIE strain) revealed some single nucleotide polymorphisms (SNPs) that only appeared in the Vietnamese strain in comparison to those of the isolates in other countries, including: (1) a change in ORF8-Nucleocapsid (accession number MT127114.1) at nucleotide (nt) 691 (**C**TC in SARS-CoV-2/NIHE/human/2020/VIE strain, **T**TC in all other isolates) but it does not change the encoded amino acid, (2) ORF3a-EM-ORF6-ORF7a region (accession number MT127115.1) has four-point changes, three of which lead to changes in the amino acid sequences, being nt 479 (**G**TA encoding Valine changed into **T**TA encoding Leucine), nt 575 (**C**GC encoding Arginine changed to **G**GC encoding Glycine) in the M gene, and nt 1126 (**G**TG encoding Valine changed to **G**A**G** encoding Glutamic acid) in ORF6. Taken together, the results provided useful information for the SARS-CoV-2 diagnostic kit and vaccine development.

Keywords: COVID-19, human coronaviruses, SARS-CoV-2, SNPs, Vietnam.

INTRODUCTION

The causative agent of the current COVID-19 pandemic (SARS-CoV-2) was first reported in Wuhan in China and has rapidly spread around the world, which resulted in a global health emergency. On March 11th, 2020, WHO publicly declared the SARS-CoV-2 outbreak to be a pandemic (Caly et al., 2020, Pachetti et al., 2020). Despite much effort to contain the virus in China, within a few months, the outbreak had reached and affected 215 countries and territories around the world. The SARS-CoV-2 is a Betacoronavirus like the SARS and MERS human coronaviruses. There are seven different strains of human coronaviruses (HCoVs) that have been detected including 229E and NL63 (Alphacoronaviruses), and OC43, HKU1, SARS, MERS, and SARS-CoV-2 (Beta coronaviruses) (Elfiky, 2020).

Coronaviruses, including the newly discovered SARS-CoV-2, are positive-sense single-stranded RNA viruses of the family Coronaviridae, causing respiratory and intestinal infections in animals and humans (Cui et al., 2019; Ahmed et al., 2020). Notably, clinical signs of SARS-CoV-2 infection are very non-specific, such as respiratory symptoms, fever, cough, shortness of breath, and pneumonia (Chen et al., 2020).

The coronavirus genome is about 30 kb long with a structure of 5′-leader-UTR - replicase - S (Spike) - E (Envelope) - M (Membrane) - N (Nucleocapsid) - 3′ UTR-poly (A), in which accessory genes are interspersed within the structural genes at the 3′ end of the genome (Fehr & Perlman, 2015). Two-thirds of the 5' end region of the genome is a highly conserved ORF1 region that encodes a polyprotein, ppa1, which is cleaved into 16 non-structural proteins required for RNA replication and transcription (Masters, 2006). The ORF1 region contains two open frames, ORF1a and ORF1b. The polyprotein translated from ORF1a/1b is involved in the ribosome's reading frame-regulating activity (Tan et al., 2005; Ziebuhr, 2004; Plant et al., 2005). In particular, this region encodes a non-structural protein that is critical for viral replication, including an RNA-dependent RNA polymerase (RdRP). The 3′ UTR also contains RNA structures required for viral RNA replication and synthesis. The spike (S) protein, the envelope (E) protein, the membrane (M) protein, and the nucleocapsid (N) protein are the main structural proteins of the virus (Tan et al., 2005; Siddell et al., 2005; Enjuanes et al., 2008; Sheikh et al., 2020). Characterization of the viral protein of SARS-CoV has shown that proteins generated from the cleavage of polyprotein pp1a/pp1ab and these structural proteins (S, M, E, and N), are similar between different coronaviruses. Studies on the evolutionary and phylogenetics of some coronaviruses such as SARS-CoV, MERS-CoV, MERS-CoV, SARS-CoV, ... have also shown the ability to use conserved gene regions such as N, E, ORF1 genes to detect different coronaviruses (Cui et al., 2019). Meanwhile, as the four major coronavirus structural proteins are the targets for diagnostic kit and vaccine development, any changes in their sequences may change the viral pathogenicity, infectivity as well as the accuracy of diagnostic assays (Chu et al,. 2020, Corman et al., 2020, Sohail et al., 2021).

According to The Vietnamese Ministry of Health (https://ncov.moh.gov.vn/trang-chu), researchers at The National Institute of Hygiene and Epidemiology have sequenced the full genome of three SARS-CoV-2 positive patients (cases #1483, #1484 and #1485) who came from South Africa on 31st Dec 2020 and revealed a new variant from South Africa. The continuous mutations of the virus may influence the effectiveness of molecular diagnostic assays. This study is carried out to better understand the molecular characteristics of the SARS-CoV-2 circulating in Vietnam and provide useful information for an effective strategy to control SARS-CoV-2 infections in Vietnam.

MATERIALS AND METHODS

SARS-CoV-2 RNA sample

A purified RNA sample of the SARS-CoV-2 strain isolated from a Vietnamese

patient (SARS-CoV-2/NIHE/human/2020/VIE) in Hanoi, Vietnam in Feb 2020 was provided by The National Institute of Hygiene and Epidemiology.

cDNA synthesis

The purified RNA sample was converted into cDNA using the 1st Strand cDNA Synthesis Kit (Thermo Scientific, USA) according to the manufacturer's instructions.

Oligonucleotide primers and nested PCR

Primers were designed based on the SARS-CoV-2 complete genome sequences in Genbank. The primers for amplification of the E, M, ORF1b, RdRp, and N genes were described in table 1. The target genes were amplified by nested PCR. 2 µL of the first PCR product was used as DNA templates for the 2nd PCR. A final reaction mixture includes PCR master mix buffer (Thermo Fisher Scientific), forward primer (10 pM), reverse primer (10 pM), DNA template (30 ng), and nuclease-free water. The PCR condition was as follows: one cycle of initial denaturation at 94 ℃ for 3 min, followed by 35 cycles of denaturation at 94 ℃ for 1 min, annealing at 55 ℃ for 40 s, and extension at 72 ℃ for 1 min. Finally, the samples were maintained at 72 ℃ for 8 min. The reactions were performed in a C1000 Thermal Cycler (Bio-Rad, USA).

Table 1. List of primers used in this study

No.	Primer name	Sequences	Target gene	PCR product length (bp)
1	E-Pn	GGAACCAATTTATGATGAACCGACG	E gene (1st PCR)	688
2	E-Mn	TGAATGACCACATGGAACGCG		
3	E-Pt	ACCAATTTATGATGAACCGACGACG	E gene (2nd PCR)	662
4	E-Mt	CGCGCAAACAGTCTGAAAGAAGC		
5	M-Pn	TGAACCGACGACGACTACTAGCG	M gene (1st PCR)	1353
6	M-Mn	AGAGGATGAAATGGTGAATTGCCC		
7	M-Pt	TTACTGCGCTTCGATTGTGTGCG	M gene (2nd PCR)	1185
8	M-Mt	ATGGTGAATTGCCCTCGTATGTTCC		
9	ORF1b-Pn	TCCAGAGTTAGTGCTAAACCACCGC	ORF1b (1st PCR)	849
10	ORF1b-Mn	ACCATCACAACCAGGCAAGTTAAGG		
11	ORF1b-Pt	AGGACTTCCTTGGAATGTAGTGCG	ORF1b (2nd PCR)	616
12	ORF1b-Mt	AGCTTTGTCACTACAAGGCTGTGC		
13	RdRp-Pn	AATCAATAGCCGCCACTAGAGG	RdRp (1st PCR)	761
14	RdRp-Mn	AGGAAGGTACACATAATCATCACCC		
15	RdRp-Pt	GGAACAAGCAAATTCTATGGTGG	RdRp (2nd PCR)	541
16	RdRp-Mt	AAGTGCTATTGAAACACACAACAGC		
17	N-Pn	ATGAGGCTGGTTCTAAATCACCC	N gene (1st PCR)	1643
18	N-Mn	TGGTGGCTCTTTCAAGTCCTCC		
19	N-Pt	TCTAAATCACCCATTCAGTACATCG	N gene (2nd PCR)	1473
20	N-Mt	AGCCCATCTGCCTTGTGTGG		

Sequence analysis

The 2nd PCR products were purified by Gel-extraction kit (Qiagen) and then cloned into plasmids using TA cloning Kit, Thermo Fisher Scientific according to the manufacturer's instructions. They were then sequenced using a BigDye terminator cycle sequencing kit (Applied Biosystems) and an automatic DNA sequencer (model 3730; Applied Biosystems). The resultant gene sequences were assembled and aligned using CLUSTAL X ver. 1.81 (Thompson et al., 1994) and MegAlign (DNAS-TAR, Madison, WI) against corresponding nucleotide sequences of SARS-CoV-2 from the GenBank database. Rooted trees were constructed using the neighbor-joining

algorithm based on the nucleotide sequences of the E, M, N, ORF6, and RdRp genes of the SARS-CoV-2/NIHE/human/2020/VIE strain and the respective regions of the 40 SARS-CoV-2 sequences in GenBank (Supplementary material 1) using the UPGMA method (Gronau and Moran 2007) in the RDP5 program (http://web.cbio.uct.ac.za/).

SNPs detection

SNPs were detected as sequence differences in multiple alignments using ClustalW (Thompson et al., 1994). Electropherograms were visually inspected using BioEdit. SNPs were identified as transitions or transversions in coding and non-coding regions by the BioEdit sequence alignment program.

RESULTS AND DISCUSSION

Gene amplification and sequencing

Five ORFs covering the M, N, E, RdRp, ORF1 genes of the SARS-CoV-2/NIHE/human/2020/VIE isolate were amplified by nested RT-PCR as described above. The PCR products were cloned into the cloning vectors then subjected to automatic DNA sequencing. The sequences were processed by Bioedit, assembled, and deposited in Genbank with accession numbers: **MT127113.1** (ORF1ab polyprotein, Exonuclease region, partial cds) **MT127114.1** (ORF8 partial cds, Nucleocapsid complete cds), **MT127115.1** (ORF3a partial cds; E, M, ORF6 complete cds; and ORF7a partial cds), **MT127116.1** (RdRp partial cds) (Table 2).

Table 2. Accession numbers of the five ORFs

Accession No.	Sequence included	Position (nt)
MT127113.1	ORF1ab polyprotein, Exonuclease region, partial cds	1–615
MT127114.1 (Length: 1411 nt)	ORF8 (partial cds)	1–128
	N (complete cds)	143–1402
MT127115.1 (Length: 1269 nt)	ORF3a (partial cds)	1–80
	E (complete cds)	33–260
	M (complete cds)	311–979
	ORF6 (complete cds)	990–1175
	ORF7a (partial cds)	1182–1269
MT127116.1 (Length: 459 nt)	RdRp (partial cds)	1–459

Abbreviations: N: Nucleoprotein; M: Membrane; E: Envelope; ORF8: open-reading frame 8; ORF3a: open-reading frames 3a; ORF6: open-reading frames 6; ORF7a: open-reading frames 7a; RdRp: RNA-dependent RNA-polymerase; nt: Nucleotide.

N gene sequence analysis

N protein is the only protein present in the nucleocapsid and has an important role in RNA synthesis and transcription (Weiwei et al., 2014; Cui et al., 2019; Sheikh et al., 2020). They consist of two separate domains, the N terminal domain (NTD) and the C terminal domain (CTD), both of which are capable of RNA binding *in vitro*, but each has different mechanisms to bind to RNA. N proteins are also strongly phosphorylated and

phosphorylation is thought to induce structural changes to increase affinity for viral RNA over non-RNA (Maache et al., 2006). N proteins are the most conserved and stable protein among CoV structural proteins (Gao et al., 2013; Sheikh et al., 2010). Studies at the nucleotide level have suggested that their mutations may accelerate Cov's evolutionary level and be closely related to the ability to adapt and infect their hosts (Weiwei et al., 2014; Cui et al., 2019; Sheikh et al., 2020; Chu et al., 2020). Moreover, it has been

shown that the N protein of SARS-CoV is not only an important B cell immunogen but also can elicit broad-based cellular immune responses and be of potential value in vaccine development for specific prophylaxis and treatment against SARS (Zhao et al., 2005). In this study, analysis of ORF8-Nucleocapsid gene region (MT127114.1), 1411 bp in length, showed that the Vietnamese isolate clustered in a separated clade with 13 other SARS-CoV-2 isolates worldwide (Fig. 1), including 3 isolates of China (MT079851.1, MT123292.2, MT079843.1) and 10 isolates of USA (MT628272.1, MT628271.1, MT628269.1, MT628268.1, MT628267.1, MT628266.1, MT628263.1, MT246667.1, MT020881.1, MT020880.1). Sequence analysis on the N gene (MT127114.1) of the SARS-CoV-2 Vietnamese isolate showed nucleotide identities of 88.53% and 66.29% compared to the N gene of SARS coronavirus Tor2 strain (**SARS-CoV**) isolated in 2004

(accession number NC_004718.3) and HCoV-EMC strain (**MERS-CoV**) isolated in 2012 (accession number NC_019843.3), respectively (data not shown). The similarity between the N gene of SARS-CoV-2 Vietnamese isolate and the SARS-CoV-2 isolates in other countries ranged from 99.7% to 99.8% (data not shown). Moreover, we found that, in the N gene region (1260 nt) of the Vietnamese isolate, there was one nucleotide change (SNP) at position 691, in which Cytosine (as in 40 other investigated isolates) changed to Thymine (C691T). A previous study has shown that Thymine and Thymidine are much less basic than Cytosine and Cytidine (Bonaccorsi et al., 1972). However, this change did not lead to amino acid substitution (Table 3) but it might contribute to the interaction between viral proteins themselves or with their host proteins. More studies need to be carried out to prove this hypothesis.
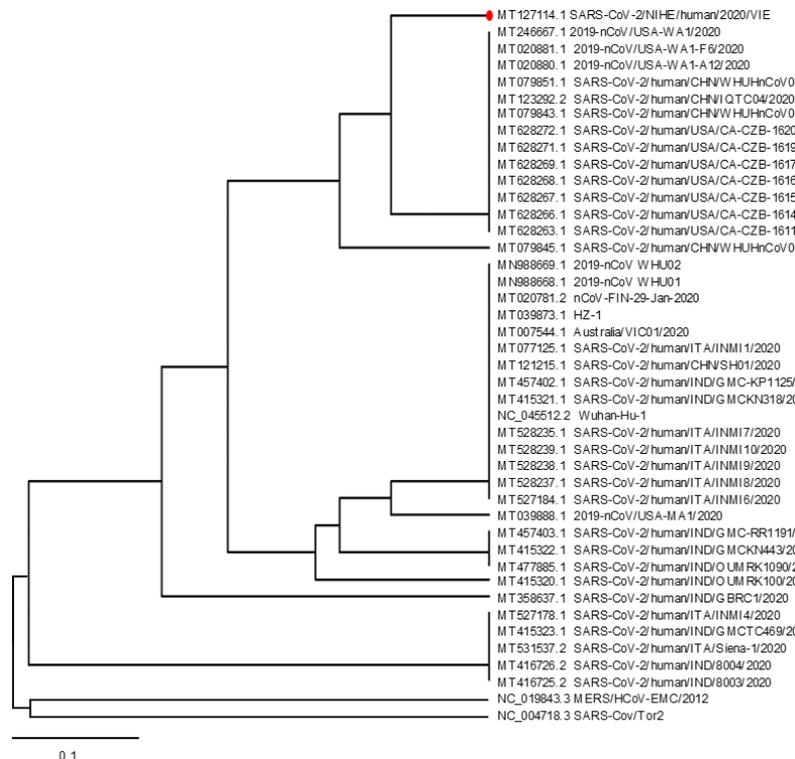
Figure 1. Phylogenetic analysis of partial genome sequences virus (SARS-CoV-2) based on sequences of the ORF8 and Nucleocapsid gene (MT127114.1) of the Vietnamese isolate and those of other isolates in different countries

Table 3. Nucleotide and amino acid changes found in the SARS-CoV-2 Vietnamese isolate

Regions	Site	Nucleotide mutations	Amino acid substitution
N (complete cds)	691	C → T	No
M (complete cds)	479517575	G → T	Val → Leu
		C → T	No
		C → G	Arg → Gly
ORF6 (complete cds)	1126	T → A	Val → Glu

Abbreviations: N: Nucleoprotein; M: Membrane; ORF6: open-reading frames 6; No: no change.

E gene sequence analysis

E protein (~ 8–12 kDa) is a small integral membrane protein found in small amounts in virions and most likely forms ion channels in the virus. This protein plays a role in virus assembly and release and other functions in the viral infectivity and pathogenicity (Tan et al., 2005; Fehr and Perlman 2015; Alsaadi & Jones 2019). The E protein is generally conserved across β-coronaviruses, especially among SARS-CoVs. In this study, the sequence of MT127115.1, 1269 nt in length,

covered 5 different ORFs arranged in the order of ORF3a-E-M-ORF6-ORF7a. The phylogenetic tree based on MT127115.1 demonstrated that the SARS-CoV-2/NIHE/human/2020/VIE belonged to one clade while the other 40 investigated isolates were located in another clade (Fig. 2). A deeper analysis of the E gene revealed that this gene of the Vietnamese isolate was not different from that of the others. In other words, we did not find any SNPs in the gene encoding E protein in the Vietnamese isolate (data not shown).
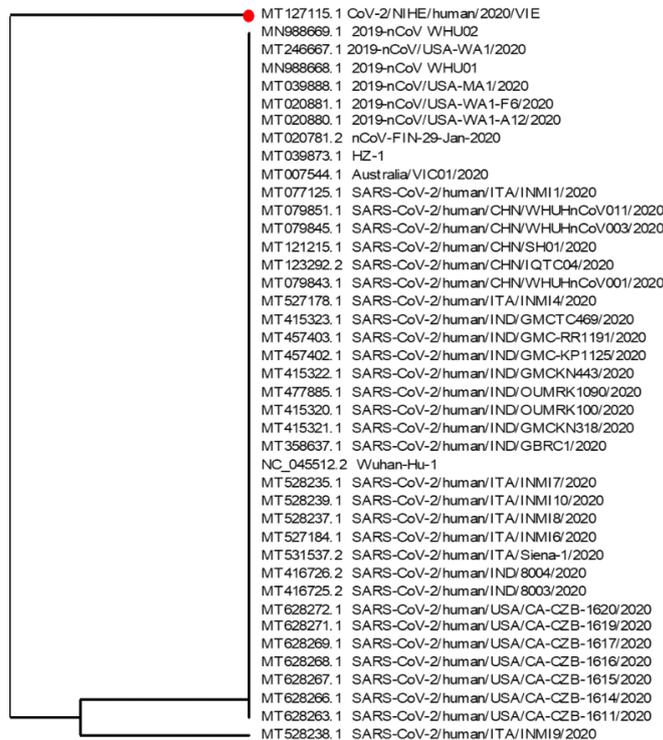
Figure 2. Phylogenetic tree of SARS-CoV-2/NIHE/human/2020/VIE virus strain generated by the UPGMA method based on nucleotide sequences of ORF3a-E-M-orf6-orf7a gene region (MT127115.1) of the Vietnamese isolate and those of other isolates worldwide.

M gene sequence analysis

M protein is the most abundant structural protein in virions with three transmembrane regions and is believed to be the main constituent of the virion shape. They have a small glycosylated N-terminal and a larger C-terminal endodomain, which is located 6−8 nm deep into the virus particle. Recent studies have shown that the M protein exists as a dimer in virions and can form two different configurations, allowing it to form the curvature of the membrane as well as to bind to nucleocapsid (Tan et al., 2005, Fehr and Perlman, 2015). The M protein is also described as the primary driver of the coronavirus budding process (Neuman et al., 2011, Alsaadi & Jones 2019). In this study, we found three variations in the M gene of the Vietnamese isolate (nt 311 - nt 979) (Fig. 4, Table 3), and two out of these three variations led to amino acid substitutions. At the nucleotide position 479, the codon **G**TA encoding Valine (Val) was changed to **T**TA encoding Leucine (Leu) in comparison to other investigated isolates. At the nucleotide position 517, the codon GCC was changed to GC**T** but it did not change the encoded amino acid; at the nucleotide position 575, the codon **C**GC encoding Arginine (Arg) was changed to **G**GC encoding Glycine (Gly).

MT 127116.1 SARS-CoV-2/NIHE/human/2020/VIE
MN988669.1 2019-nCoV WHU02
MT 246667.1 2019-nCoV/USA-WA1/2020
MN988668.1 2019-nCoV WHU01
MT 039888.1 2019-nCoV/USA-MA1/2020
MT 020881.1 2019-nCoV/USA-WA1-F6/2020
MT 020880.1 2019-nCoV/USA-WA1-A12/2020
MT 020781.2 nCoV-FIN-29-Jan-2020
MT 039873.1 HZ-1
MT 007544.1 Australia/VIC01/2020
MT 077125.1 SARS-CoV-2/human/ITA/INMI1/2020
MT 079851.1 SARS-CoV-2/human/CHN/WHUHnCoV011/2020
MT 079845.1 SARS-CoV-2/human/CHN/WHUHnCoV003/2020
MT 121215.1 SARS-CoV-2/human/CHN/SH01/2020
MT 123292.2 SARS-CoV-2/human/CHN/IQTC04/2020
MT 079843.1 SARS-CoV-2/human/CHN/WHUHnCoV001/2020
MT 527178.1 SARS-CoV-2/human/ITA/INMI4/2020
MT 415323.1 SARS-CoV-2/human/IND/GMCTC469/2020
MT 457403.1 SARS-CoV-2/human/IND/GMC-RR1191/2020
MT 457402.1 SARS-CoV-2/human/IND/GMC-KP1125/2020
MT 415322.1 SARS-CoV-2/human/IND/GMCKN443/2020
MT 477885.1 SARS-CoV-2/human/IND/OUMRK1090/2020
MT 415321.1 SARS-CoV-2/human/IND/GMCKN318/2020
MT 358637.1 SARS-CoV-2/human/IND/GBRC1/2020
NC_045512.2 Wuhan-Hu-1
MT 528235.1 SARS-CoV-2/human/ITA/INMI7/2020
MT 528239.1 SARS-CoV-2/human/ITA/INMI10/2020
MT 528238.1 SARS-CoV-2/human/ITA/INMI9/2020
MT 528237.1 SARS-CoV-2/human/ITA/INMI8/2020
MT 527184.1 SARS-CoV-2/human/ITA/INMI6/2020
MT 531537.2 SARS-CoV-2/human/ITA/Siena-1/2020
MT 416726.2 SARS-CoV-2/human/IND/8004/2020
MT 416725.2 SARS-CoV-2/human/IND/8003/2020
MT 628272.1 SARS-CoV-2/human/USA/CA-CZB-1620/2020
MT 628271.1 SARS-CoV-2/human/USA/CA-CZB-1619/2020
MT 628269.1 SARS-CoV-2/human/USA/CA-CZB-1617/2020
MT 628268.1 SARS-CoV-2/human/USA/CA-CZB-1616/2020
MT 628267.1 SARS-CoV-2/human/USA/CA-CZB-1615/2020
MT 628266.1 SARS-CoV-2/human/USA/CA-CZB-1614/2020
MT 628263.1 SARS-CoV-2/human/USA/CA-CZB-1611/2020
MT 415320.1 SARS-CoV-2/human/IND/OUMRK100/2020

Figure 3. Phylogenetic tree of partial genome sequences virus (SARS-CoV-2) based on nucleotide sequences of similar region of the Vietnamese isolate (RdRp gene, MT127116.1) and those of other isolates worldwide
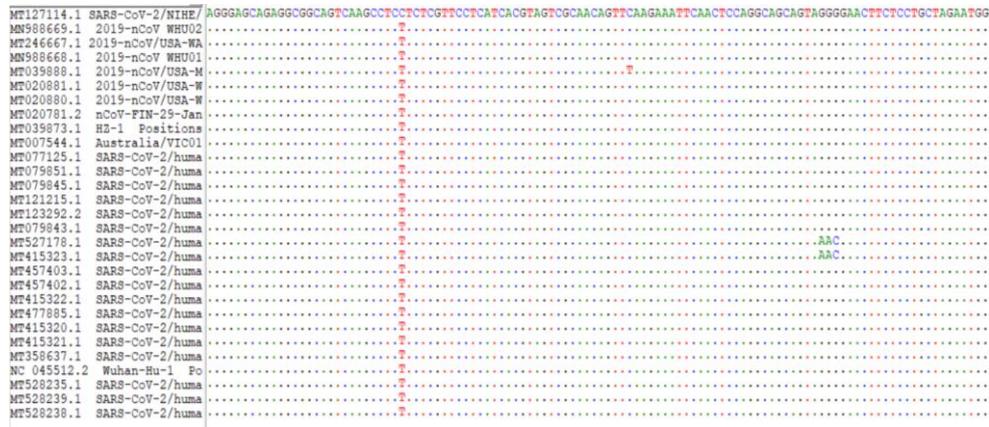
A

B

C

Figure 4. Clustal alignment of the nucleotide sequences of the SARS-CoV-2/Human/NIHE/ 2020/VIE with 40 other SARS-CoV-2 isolates. Alignment of the nucleotide sequences of the N gene (A), M gene (B), and ORF6 gene (C); CoV-2/NIHE/: SARS-CoV-2/Human/NIHE/2020/VIE

RdRp and ORF6 sequence analysis

The RdRp protein (~ 240 to 450 kDa) contains a catalytic core with a clear resemblance to the human right hand with palm, fingers, and thumb domains (Hillen et al., 2020; Aftab et al., 2020); RdRp is considered to be a conserved protein within RNA viruses and plays a crucial role in the viral life cycle including viral gene transcription and replication in concert with other viral and host factors (Fehr and Perlman, 2015). This region is targeted for the inhibition of viral replication as an effective therapeutic approach (Aftab et al., 2020). Any alternatives in this region may affect the efficiency of antiviral drug treatment. In this study, we found that there were no differences between the nucleotide sequence of RdRp of the Vietnamese isolate (MT127115.1, 459 nt) and other isolates with exception of the Indian isolate (MT415320.1) (Fig. 3).

The ORF6 of SARS-CoV-2 has the strongest suppression on both primary interferon production and interferon signaling (Miorin et al., 2020). ORF6-deleted SARS-CoV-2 may be considered for the development of an intranasal live-but-attenuated vaccine against COVID-19 (Yuen et al., 2020). In this study, the ORF6 region of the Vietnamese isolate had one SNP at position 1126 in the codon **G**T**G** encoding Valine (Val) changed to **GA**G encoding Glutamic acid (Glu) compared to other isolates (Table 3). The SNP found in ORF6 of the Vietnamese isolate suggested that it may influence the primary interferon production and interferon signaling induced by the virus infection; however, more studies need to be carried out to demonstrate it.

CONCLUSION

In this study, several SNPs in the genome of the Vietnamese SARS-CoV-2 strain isolated from Hanoi city were detected and some of them led to amino acid substitutions. Our results reinforced that SARS-CoV-2 can acquire novel variations rapidly as it spreads in various human populations and suggested that combining genomic data with epidemiological data has the potential to inform public health intervention policy. Besides, our data provided useful information for diagnostic kit and vaccine development to control SARS-CoV-2 infections in Vietnam.

REFERENCES

Aftab, S. O., Ghouri, M. Z., Masood, M. U., Haider, Z., Khan, Z., Ahmad, A., & Munawar, N., 2020. Analysis of SARS-CoV-2 RNA-dependent RNA polymerase as a potential therapeutic drug target using a computational approach. *Journal of Translational Medicine*, 18(1), 1−15.

Ahmed S. F., Quadeer A. A., McKay M. R., 2020. Preliminary identification of potential vaccine targets for the COVID-19 coronavirus (SARS-CoV-2) based on SARS-CoV immunological studies. *Viruses*, 12(3): 254.

Alsaadi E. A., Jones I. M., 2019. Membrane binding proteins of coronaviruses. *Future Virology*, 14(4): 275−286.

Bonaccorsi R., Pullman A., Scrocco E., Tomasi J., 1972. The molecular electrostatic potentials for the nucleic acid bases: adenine, thymine, and cytosine. *Theoretica Chimica Acta*, 24(1): 51−60.

Caly L., Druce J. D., Catton M. G., Jans D. A., Wagstaff K. M., 2020. The FDA-approved drug ivermectin inhibits the replication of SARS-CoV-2 in vitro. *Antiviral Research.*, https://doi.org/10.1016/j.antiviral.2020.104787

Chen L., Liu W., Zhang Q., Xu K., Ye G., Wu W., Liu Y., 2020. RNA based mNGS approach identifies a novel human coronavirus from two individual pneumonia cases in 2019 Wuhan outbreak. *Emerging Microbes Infections*, 9(1): 313−319.

Chu D. K., Pan Y., Cheng S. M., Hui K. P., Krishnan P., Liu Y., Poon L. L., 2020.

Molecular diagnosis of a novel coronavirus (2019-nCoV) causing an outbreak of pneumonia. *Clinical Chemistry*, 66(4): 549−555.

Corman V. M., Landt O., Kaiser M., Molenkamp R., Meijer A., Chu D. K., ..., Drosten C., 2020. Detection of 2019 novel coronavirus (2019-nCoV) by real-time RT-PCR. *Eurosurveillance.*, 25(3): 2000045.

Cui J., Li F., Shi Z. L., 2019. Origin and evolution of pathogenic coronaviruses. *Nature Reviews Microbiology*, 17(3): 181−192.

Elfiky A. A., 2020. SARS-CoV-2 RNA dependent RNA polymerase (RdRp) targeting: An in silico perspective. *Journal of Biomolecular Structure and Dynamics*. https://doi.org/ 10.1080/07391102. 2020.1761882

Enjuanes L., DeDiego M. L., Álvarez E., Deming D., Sheahan T., Baric R., 2008. Vaccines to prevent severe acute respiratory syndrome coronavirus-induced disease. *Virus Research*, 133(1): 45−62.

Fehr A. R., Perlman S., 2015. Coronaviruses: an overview of their replication and pathogenesis. *Coronaviruses*. https://doi.org/10.1007/978-1-4939-2438-7_1

Gao J., Lu G., Qi J., Li Y., Wu Y., Deng Y., Gao G. F., 2013. Structure of the fusion core and inhibition of fusion by a heptad repeat peptide derived from the S protein of Middle East respiratory syndrome coronavirus. *Journal of Virology*, 87(24): 13134−13140.

Gronau I., Moran S., 2007. Optimal implementations of UPGMA and other common clustering algorithms. *Information Processing Letters*, 104(6): 205−210.

Hillen, H. S., Kokic, G., Farnung, L., Dienemann, C., Tegunov, D., & Cramer, P., 2020. Structure of replicating SARS-CoV-2 polymerase. *Nature*, 584(7819): 154−156.

Maache M., Komurian F., Rajoharison A., Perret M., Berland J. L., Pouzol S., Paranhos G., 2006. False-positive results in a recombinant severe acute respiratory syndrome-associated coronavirus (SARS-CoV) nucleocapsid-based western blot assay were rectified by the use of two subunits (S1 and S2) of spike for detection of antibody to SARS-CoV. *Clinical and Vaccine Immunology*, 13(3): 409−414.

Masters P. S., 2006. The molecular biology of coronaviruses. *Advances in Virus Research*, 66: 193−292.

Miorin, L., Kehrer, T., Sanchez-Aparicio, M. T., Zhang, K., Cohen, P., Patel, R. S., ... & García-Sastre, A., 2020. SARS-CoV-2 Orf6 hijacks Nup98 to block STAT nuclear import and antagonize interferon signaling. *Proceedings of the National Academy of Sciences*, 117(45): 28344−28354.

Neuman B. W., Kiss G., Kunding A. H., Bhella D., Baksh M. F., Connelly S., Buchmeier, M. J., 2011. A structural analysis of M protein in coronavirus assembly and morphology. *Journal of Structural Biology*, 174(1): 11−22.

Pachetti M., Marini B., Benedetti F., Giudici F., Mauro E., Storici, P., Ippodrino R., 2020. Emerging SARS-CoV-2 mutation hot spots include a novel RNA-dependent-RNA polymerase variant. *Journal of Translational Medicine*, 18: 1−9.

Plant E. P., Pérez G. C., Jacobs J. L., Mukhopadhyay B., Hennig M., Dinman J. D., 2005. A three-stemmed mRNA pseudoknot in the SARS coronavirus frameshift signal. *PLoS Biol.*, 3(6): e172.

Sheikh A., Al-Taher A., Al-Nazawi M., Al-Mubarak A. I., Kandeel M., 2020. Analysis of preferred codon usage in the coronavirus N genes and their implications for genome evolution and vaccine design. *Journal of Virological Methods*, 277: 113806. https://doi.org/ 10.1016/j.jviromet.2019.113806

Siddell S. G., Ziebuhr J., Snijder E. J., 2005. Coronaviruses, toroviruses, and arteriviruses *Topley and Wilson's Microbiology and Microbial Infections*., pp. 823−856.

Sohail M. S., Ahmed S. F., Quadeer A. A., McKay M., 2020. In silico T cell epitope identification for SARS-CoV-2: Progress and perspectives. *Available at SSRN*., http://dx.doi.org/10.2139/ssrn.3720371

Tan Y. J., Lim S. G., Hong W., 2005. Characterization of viral proteins encoded by the SARS-coronavirus genome. *Antiviral Research.*, 65(2): 69−78.

Thompson J. D., Higgins D. G., Gibson T. J., 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research.*, 22(22): 4673−4680.

Weiwei H., Qinghua Y., Liqi Z., Haofei L., Shanshan Z., Qi G., Qian Y., 2014. Complete genomic sequence of the coronavirus transmissible gastroenteritis virus SHXB isolated in China. *Archives of Virology*, 159(9): 2295−2302.

Yuen C. K., Lam J. Y., Wong W. M., Mak L. F., Wang X., Chu H., Kok K. H., 2020. SARS-CoV-2 nsp13, nsp14, nsp15 and orf6 function as potent interferon antagonists. *Emerging Microbes Infections*, 9(1): 1418−1428.

Zhao P., Cao J., Zhao L. J., Qin Z. L., Ke J. S., Pan W., Qi Z. T., 2005. Immune responses against SARS-coronavirus nucleocapsid protein induced by DNA vaccine. *Virology*, 331(1): 128−135.

Ziebuhr J., 2004. Molecular biology of severe acute respiratory syndrome coronavirus. *Current Opinion in Microbiology*, 7(4): 412−419.