medRxiv preprint doi: https://doi.org/10.1101/2021.03.18.21253734; this version posted March 24, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

# SARS-CoV-2 spike protein gene variants with N501T and G142D mutation dominated infections in minks in the US

Hugh Y. Cai[1]*, Allison Cai[2]

1. Animal Health Laboratory, University of Guelph, Guelph, Ontario, Canada

2. Faculty of Science, University of British Colombia, BC, Canada

*Corresponding author: hcai@uoguelph.ca

## Abstract

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) have infected human and animals world wide. Since the first SARS-CoV-2 sequences identified from samples collected in December 2019, the virus has been evolving. Now, in the GISAID database of SARS-CoV-2 genomes as of March 17, 2021, 95.7% (758,684 of 792,388) of the SARS-CoV-2 genome has the D614G mutation in the spike protein gene; and 22% (178,186 of 792.388) had N501Y mutation which occurred in the more virulent UK variants, South Africa variant and Brazil variant. Large number of minks were infected with the virus which containing the Y453F mutation in Europe, causing zoonosis concerns. To evaluate the genetic characteristics of the US mink-derived SARS-CoV-2 sequences, we analyzed all animal-derived (977), all Canadian (19,529) and US (173,277) SARS-CoV-2 sequences deposited in GISAID from December 2019 to March 12, 2021, and identified 2 dominant novel variants, the N501T-G142D variant and N501T-G142D-F486L variant, in the US mink-derived SARS-CoV-2 sequences. These variants were not found in minks from Canada or other countries. The Y453F mutation was not identified in the mink-derive sequences in the US and Canada. The N501T mutation occurred two months earlier in the human than in the minks in the US. The results of our study highlighted the importance of on going monitoring the mutations of the SARS-CoV-2 virus in animals.

## Keywords

SARS-CoV-2, COVID-19, spike protein, S gene, mutation, variant, mink

## Introduction

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2021.03.18.21253734; this version posted March 24, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Genomic surveillance of circulating SARS-CoV-2 variants is critical for epidemiologic tracking and infection control. Rapid global spread of a variant harbouring a mutation located on the SARS-CoV-2 spike protein (S protein) has been reported (Phan 2020; Na, et al., 2021). In particular, codon N501 has been found to be critical; the UK variant (lineage B.1.1.7) with an N501Y mutation increased transmissibility by 50% and is potentially more virulent (ECDC, 2021); a similar phenomenon was observed with the South African variant lineage B.1.351 (Kai, 2021) and the Brazil variant lineage B.1.525 (Voloch, 2020), both of which have the N501Y mutation. Large numbers of mutants with Y453F mutations were found in infected mink in Denmark, which might have been spread from human to animal and back to human (Munnik et al., 2021). We analyzed the sequences of the SARS-CoV-2 S protein gene collected from mink in the USA and Canada, and identified 2 dominant novel variants with 2 and 3 mutations each.

## Materials and Methods

From the GISAID (Global Initiative on Sharing All Influenza Data, https://www.gisaid.org), we downloaded the following SARS-CoV-2 sequences collected from December 2019 to the date of writing (March 12, 2021): all animal-derived sequences (977), all sequences collected in Canada (19,529) and the US (173,277). The SARS-CoV-2 isolate Wuhan-Hu-1 collected in December 19, 2019 (GenBank accession NC045512) was used as a reference for mutation analysis. All nucleotide position labeling in our study was based on alignment with this sequence. SARS-CoV-2 full genome sequences were multiple aligned with the bioinformatic software, Geneious v.11 (Auckland, New Zealand) using “Map to a Reference Assembly” function. The aligned sequences were visually examined to confirm that they were aligned properly. The variants/SNP were identified by the software automatically and verified by visual confirmation. Short fragments (30 nt) containing the novel mutations identified in our study were used as queries to blast search the local databases consisting of the downloaded sequences to determine the incidence of the novel mutations using the “BLAST” function of the Geneious software. The global incidence of “N501Y” was determined using the “search” function with substitution “N501Y”. The global incidence of “N501T” was not determined since the pre-set substitution “N501T” not available in GISAID.

## Results and Discussion

medRxiv preprint doi: https://doi.org/10.1101/2021.03.18.21253734; this version posted March 24, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

The N501Y mutation that was common to UK, South African, and Brazil variants, and the Y453F mutation that widely prevailed in COVID-19 positive mink in Denmark and the Netherlands were absent in all the US and Canadian mink SARS-CoV-2 sequences examined. Interestingly, an N501T mutation was found in almost all US mink SARS-CoV-2 sequences (100 of 101; 99%); furthermore, all N501T mutant sequences, except 2 of poor quality, had the G142D mutation (Supplementary Table 1). Based on GISAID records, all US mink SARS-CoV-2 sequences were collected from Michigan and Wisconsin in October, 2020. We found that, except 1 sequences with poor quality, all (54) Michigan mink sequences had the F486L mutation in addition to the N501T and G142D mutations (Supplementary Table 1). Hereafter, we designated the sequences with N501T and G142D, but without F486L, mutations as the N501T-G142D variant; and the ones with all 3 novel mutations as the N501T-G142D-F486L variant. In addition to the above 3 mutations, G614D, the mutation that dominated the SARS-CoV-2 sequences, existed in all US and Canadian mink-derived sequences (Supplementary Table 1, Supplementary Table 2). No N501T-G142D nor N501T-G142D-F486L variants were found in mink from Canada or other countries. Outside the US, 5 mink sequences collected from April to June 2020 in Netherlands had the N501T mutation (without G142D) (Supplementary Table 3), which is consistent with a previous report on N501T (Elaswad et al., 2020); however, different from the study, we did not find F486L mutations in the Netherlands mink-derived sequences.

In order to determine the origin of the 2 novel variants, we analyzed all human-derived SARS-CoV-2 sequences collected from the US and Canada. Three N501T-G142D variant and 2 N501T-G142D-F486L variant sequences were identified from the sequences collected in early October 2020 from Wisconsin and Michigan, respectively. Furthermore, we identified N501T mutants without G142D and F486L in US human-derived sequences collected as early as August 2020, which was earlier than the mink-derived sequences collected in October, 2020 (Supplementary Table 4, Fig. 1). In Canada, the N501T mutation only existed in human-derived sequences (Supplementary Table 2). The N501T mutation was described in 2 Italian human-derived sequences collected in August 2020 (Fiorntini et al, 2021). We therefore speculate that the N501T mutation may have evolved during human infection and was then transmitted to mink populations in the US, and it is less likely that the mutations occurred in mink population after infected with human strains of the virus.

medRxiv preprint doi: https://doi.org/10.1101/2021.03.18.21253734; this version posted March 24, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

To evaluate the transmissibility of N501T mutants, the incidence of N501T and N501Y mutants was enumerated from US and Canadian GISAID SARS-CoV-2 sequences collected up to March 12, 2021: 2.4% (4,097 of 173,277) and 0.77% (1,339 of 173,277) of US sequences had the N501Y and N501T maturation, respectively; and 1.6% (315/19,529) and 0.06% (12/19,529) of Canadian sequences had the N501Y and the N501T mutation, respectively (Supplementary Table 5). Comparatively, the incidence of N501Y is much lower in the US and Canada than globally. Our analysis results indicate that the N501T mutants had weaker transmissibility compared to the N501Y mutants that include UK B.1.17, South Africa B.1.351, and Brazil P.1. However, it cannot be ruled out that the different time of introduction of N501Y and N501T (April 2020 vs August 2020) may have affected the occurrences of the 2 mutants (Fig. 1).

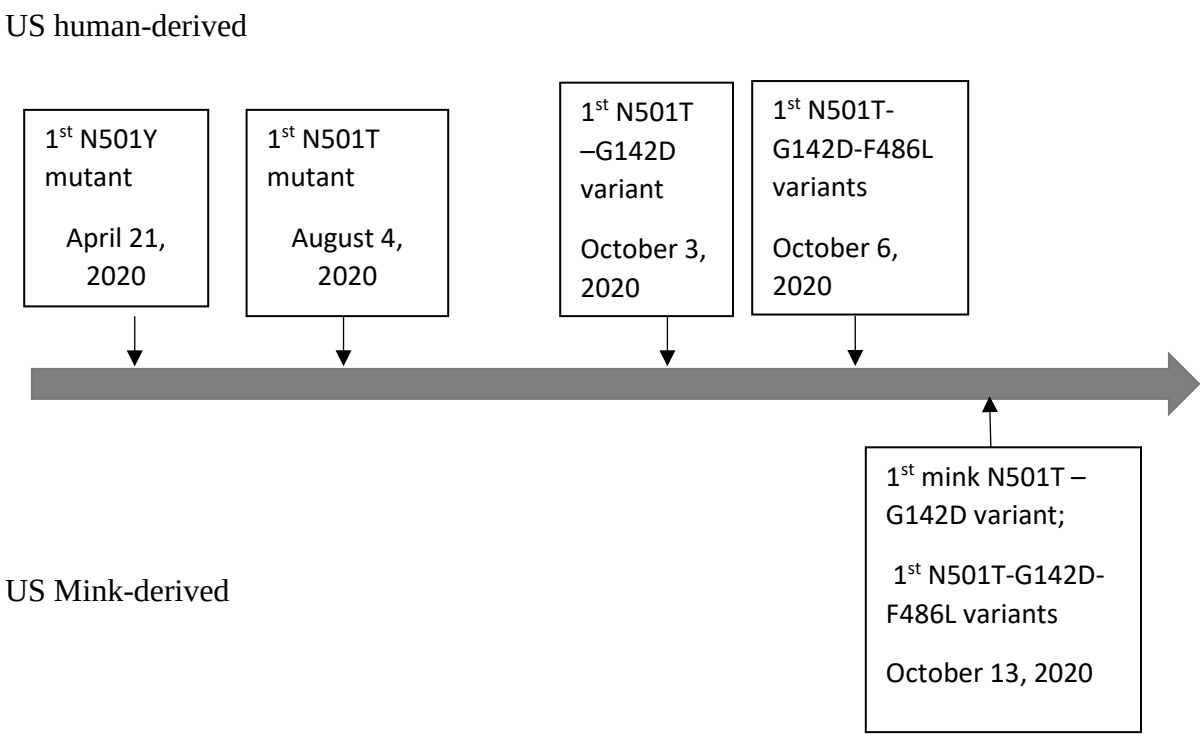


Fig. 1. Emerging timeline of human-derived (above the timeline) and mink-derived (below the timeline) SARS-CoV-2 spike protein N501T-G142D and N501T-G142D-F486L variants in the US. Scale is not in proportion.

medRxiv preprint doi: https://doi.org/10.1101/2021.03.18.21253734; this version posted March 24, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

It has been described that the mutations at S protein N501 can significantly increase the transmissibility, and likely virulence of SARS-CoV-2; the UK variant N501Y is an example. It was predicted that a single N501T mutation might significantly enhance the binding affinity between 2019-nCoV RBD and human ACE2 (Wan et al., 2020). Our analysis revealed that the N501T and G142DL mutations occurred in 99% of mink-derived sequences collected in the US; all, except one with poor quality, mink sequences collected from Michigan also contained the F486L mutation. The large number of new variants occurring in the mink population warrants further study on how these changes affected their interaction with the ACE2 receptor, and thereby the transmissibility, virulence, and immunogenicity in humans and mink. It is important to monitor the emerging new variants and determine their impact on human and animal health.

**Declarations:**

**Ethics approval and consent to participate**

Not applicable.

**Consent for publication**

All authors have read and approved the final version for publication.

**Availability of data and materials**

All sequence data used in this study were available from the GISAID's hCov-19 Database. GISAID sequences with N501T, G142D and F486L mutations are listed in Supplementary Tables.

**Competing interests**

The authors declare no competing interests.

**Funding Support:**

This study was supported the Animal Health Laboratory (AHL), University of Guelph, Canada

medRxiv preprint doi: https://doi.org/10.1101/2021.03.18.21253734; this version posted March 24, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

## Authors' contributions

HYC conceived the study. HYC and AC collected and analyzed, interpret the data and wrote the article.

## Acknowledgement

We gratefully acknowledge the authors, originating and submitting laboratories of the sequences from GISAID's hCov-19 Database on which this research is based. See supplement acknowledgement for details.

## References

ECDC (21 January 2021). "Risk related to the spread of new SARS-CoV-2 variants of concern in the EU/EEA - first update". European Centre for Disease Prevention and Control. Retrieved 2 February 2021.

Elaswad A, Fawzy M, Basiouni S, and ShehataAA. 2020. Mutational spectra of SARS-CoV-2 isolated from animals PeerJ. 8: e10609. Published online 2020 Dec 18. doi: 10.7717/peerj.10609

Fiorentini S et al. 2021. First detection of SARS-CoV-2 spike protein N501 mutation in Italy in August, 2020. Lancet Infect Dis 2021 Published Online, January 12, 2021, https://doi.org/10.1016/S1473-3099(21)00007-4 (Retrieved March 12, 2021)

Kupferschmidt K. 2021. New mutations raise specter of 'immune escape. Science. 371 (6527): 329–330. doi:10.1126/science.371.6527.329. PMID 33479129. Retrieved 25 January 2021.

Munnink BBO, Sikkema1 RS, Nieuwenhuijse DF et al., 2021. Transmission of SARS-CoV-2 on mink farms between humans and mink and back to humans. Science. Vol. 371, Issue 6525, pp. 172-177 DOI: 10.1126/science.abe5901

medRxiv preprint doi: https://doi.org/10.1101/2021.03.18.21253734; this version posted March 24, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Na, W, Moon, H and Song, D A. 2021. comprehensive review of SARS-CoV-2 genetic mutations and lessons from animal coronavirus recombination in one health perspective. J Microbiol. 59, 332–340 https://doi.org/10.1007/s12275-021-0660-4

Phan, T. 2020 Genetic diversity and evolution of SARS-CoV-2. Infect Genet Evol. 81:104260

Voloch, CM et al.2020. Genomic characterization of a novel SARS-CoV-2 lineage from Rio de Janeiro, Brazil . Retrieved 15 January 2021. doi:10.1101/2020.12.23.20248598 – via medRxiv.

Wan Y, Shang J, Graham R, Baric RS, Li F. 2020. Receptor recognition by the novel coronavirus from Wuhan: an analysis based on decade-long structural studies of SARS coronavirus. J Virol 94:e00127-20. https://doi.org/10.1128/JVI.00127-20.