

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

3 Hugh Y. Cai^{1*}, Allison Cai²

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

5 2. Faculty of Science, University of British Columbia, BC, Canada

6

7 *Corresponding author: hcai@uoguelph.ca

8 **Abstract**

9 Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) have infected human and
10 animals world wide. Since the first SARS-CoV-2 sequences identified from samples collected in
11 December 2019, the virus has been evolving. Now, in the GISAID database of SARS-CoV-2
12 genomes as of March 17, 2021, 95.7% (758,684 of 792,388) of the SARS-CoV-2 genome has
13 the D614G mutation in the spike protein gene; and 22% (178,186 of 792.388) had N501Y
14 mutation which occurred in the more virulent UK variants, South Africa variant and Brazil
15 variant. Large number of minks were infected with the virus which containing the Y453F
16 mutation in Europe, causing zoonosis concerns. To evaluate the genetic characteristics of the US
17 mink-derived SARS-CoV-2 sequences, we analyzed all animal-derived (977), all Canadian
18 (19,529) and US (173,277) SARS-CoV-2 sequences deposited in GISAID from December 2019
19 to March 12, 2021, and identified 2 dominant novel variants, the N501T-G142D variant and
20 N501T-G142D-F486L variant, in the US mink-derived SARS-CoV-2 sequences. These variants
21 were not found in minks from Canada or other countries. The Y453F mutation was not identified
22 in the mink-derive sequences in the US and Canada. The N501T mutation occurred two months
23 earlier in the human than in the minks in the US. The results of our study highlighted the
24 importance of on going monitoring the mutations of the SARS-CoV-2 virus in animals.

25 **Keywords**

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

27 **Introduction**

28 Genomic surveillance of circulating SARS-CoV-2 variants is critical for epidemiologic tracking
29 and infection control. Rapid global spread of a variant harbouring a mutation located on the
30 SARS-CoV-2 spike protein (S protein) has been reported (Phan 2020; Na, et al., 2021). In
31 particular, codon N501 has been found to be critical; the UK variant (lineage B.1.1.7) with an
32 N501Y mutation increased transmissibility by 50% and is potentially more virulent (ECDC,
33 2021); a similar phenomenon was observed with the South African variant lineage B.1.351 (Kai,
34 2021) and the Brazil variant lineage B.1.525 (Voloch, 2020), both of which have the N501Y
35 mutation. Large numbers of mutants with Y453F mutations were found in infected mink in
36 Denmark, which might have been spread from human to animal and back to human (Munnik et
37 al., 2021). We analyzed the sequences of the SARS-CoV-2 S protein gene collected from mink
38 in the USA and Canada, and identified 2 dominant novel variants with 2 and 3 mutations each.

39 **Materials and Methods**

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

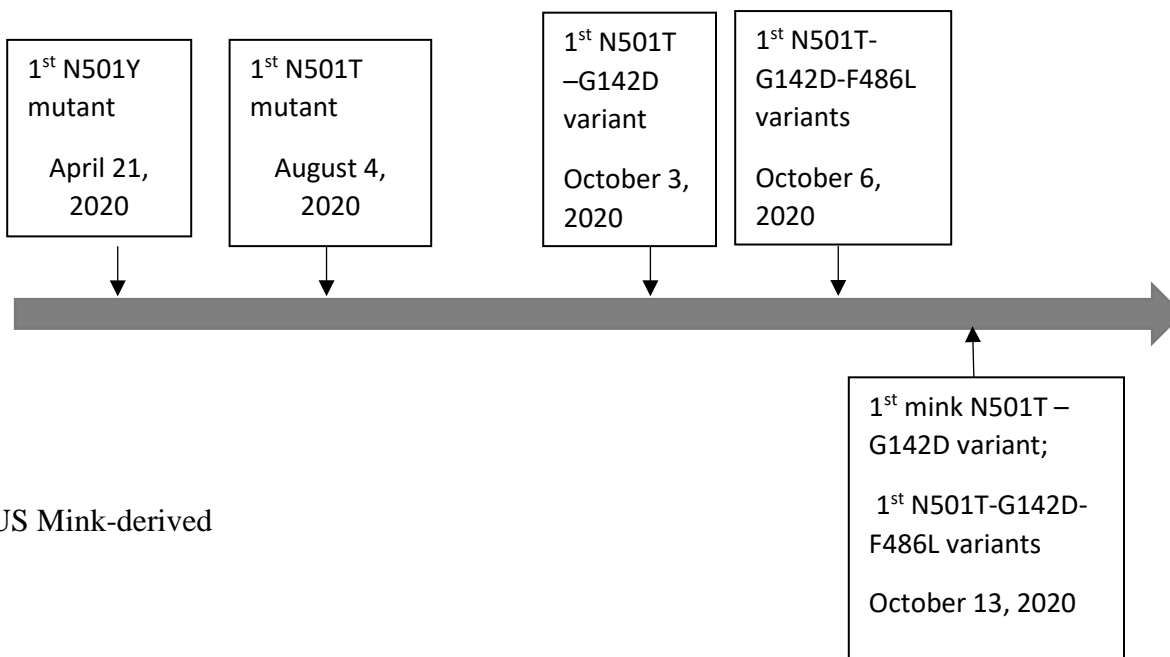
56 **Results and Discussion**

57 The N501Y mutation that was common to UK, South African, and Brazil variants, and the
58 Y453F mutation that widely prevailed in COVID-19 positive mink in Denmark and the
59 Netherlands were absent in all the US and Canadian mink SARS-CoV-2 sequences examined.
60 Interestingly, an N501T mutation was found in almost all US mink SARS-CoV-2 sequences (100
61 of 101; 99%); furthermore, all N501T mutant sequences, except 2 of poor quality, had the
62 G142D mutation (Supplementary Table 1). Based on GISAID records, all US mink SARS-CoV-
63 2 sequences were collected from Michigan and Wisconsin in October, 2020. We found that,
64 except 1 sequences with poor quality, all (54) Michigan mink sequences had the F486L mutation
65 in addition to the N501T and G142D mutations (Supplementary Table 1). Hereafter, we
66 designated the sequences with N501T and G142D, but without F486L, mutations as the N501T-
67 G142D variant; and the ones with all 3 novel mutations as the N501T-G142D-F486L variant. In
68 addition to the above 3 mutations, G614D, the mutation that dominated the SARS-CoV-2
69 sequences, existed in all US and Canadian mink-derived sequences (Supplementary Table 1,
70 Supplementary Table 2). No N501T-G142D nor N501T-G142D-F486L variants were found in
71 mink from Canada or other countries. Outside the US, 5 mink sequences collected from April to
72 June 2020 in Netherlands had the N501T mutation (without G142D) (Supplementary Table 3),
73 which is consistent with a previous report on N501T (Elaswad et al., 2020); however, different
74 from the study, we did not find F486L mutations in the Netherlands mink-derived sequences.

75 In order to determine the origin of the 2 novel variants, we analyzed all human-derived
76 SARS-CoV-2 sequences collected from the US and Canada. Three N501T-G142D variant and 2
77 N501T-G142D-F486L variant sequences were identified from the sequences collected in early
78 October 2020 from Wisconsin and Michigan, respectively. Furthermore, we identified N501T
79 mutants without G142D and F486L in US human-derived sequences collected as early as August
80 2020, which was earlier than the mink-derived sequences collected in October, 2020
81 (Supplementary Table 4, Fig. 1). In Canada, the N501T mutation only existed in human-derived
82 sequences (Supplementary Table 2). The N501T mutation was described in 2 Italian human-
83 derived sequences collected in August 2020 (Fiorntini et al, 2021). We therefore speculate that
84 the N501T mutation may have evolved during human infection and was then transmitted to mink
85 populations in the US, and it is less likely that the mutations occurred in mink population after
86 infected with human strains of the virus.

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

US human-derived



US Mink-derived

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

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

111

112

113 **Declarations:**

114 **Ethics approval and consent to participate**

115 Not applicable.

116 **Consent for publication**

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

118 **Availability of data and materials**

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

122 **Competing interests**

123 The authors declare no competing interests.

124 **Funding Support:**

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

126 **Authors' contributions**

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

129 **Acknowledgement**

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

133

134 **References**

135

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

139

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

142

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

146

147 Kupferschmidt K. 2021. New mutations raise specter of 'immune escape. Science. 371 (6527):
148 329–330. doi:10.1126/science.371.6527.329. PMID 33479129. Retrieved 25 January 2021.

149

150 Munnink BBO, Sikkema1 RS, Nieuwenhuijse DF et al., 2021. Transmission of SARS-CoV-2 on
151 mink farms between humans and mink and back to humans. Science. Vol. 371, Issue 6525, pp.
152 172-177 DOI: 10.1126/science.abe5901

153

- 154 Na, W, Moon, H and Song, D A. 2021. comprehensive review of SARS-CoV-2 genetic
155 mutations and lessons from animal coronavirus recombination in one health perspective. J
156 Microbiol. 59, 332–340 <https://doi.org/10.1007/s12275-021-0660-4>
157
- 158 Phan, T. 2020 Genetic diversity and evolution of SARS-CoV-2. Infect Genet Evol. 81:104260
159
- 160 Voloch, CM et al.2020. Genomic characterization of a novel SARS-CoV-2 lineage from Rio de
161 Janeiro, Brazil . Retrieved 15 January 2021. doi:10.1101/2020.12.23.20248598 – via medRxiv.
162
- 163 Wan Y, Shang J, Graham R, Baric RS, Li F. 2020. Receptor recognition by the novel
164 coronavirus from Wuhan: an analysis based on decade-long structural studies of SARS
165 coronavirus. J Virol 94:e00127-20. <https://doi.org/10.1128/JVI.00127-20>.