Emergence and spread of SARS-CoV-2 variants from farmed mink to humans and back during the epidemic in Denmark, June-November 2020.

Thomas Bruun Rasmussen[1*], Amanda Gammelby Qvesel[1,2,3], Anders Gorm Pedersen[2,3], Ann Sofie Olesen[1], Jannik Fonager[1], Morten Rasmussen[1], Raphael Niklaus Sieber[4], Marc Stegger[4], Francisco Fernando Calvo Artavia[5], Esben Rahbek Thuesen[2,3], Marlies Jilles Francine Goedknegt[2], Louise Lohse[1], Sten Mortensen[5], Anders Fomsgaard[1], Anette Boklund[6], Anette Bøtner[6], Graham J. Belsham[6*]

1: Department of Virus & Microbiological Special Diagnostics, Statens Serum Institut, Artillerivej 5, DK-2300 Copenhagen S, Denmark

2: Department of Health Technology, Section for Bioinformatics, Technical University of Denmark, DK-2800 Kgs. Lyngby, Denmark

3: PandemiX Center, Department of Science and Environment, Roskilde University, DK-4000 Roskilde, Denmark

4: Department of Bacteria, Parasites and Fungi, Statens Serum Institut, Artillerivej 5, DK-2300 Copenhagen S, Denmark

5: Danish Veterinary and Food Administration, Ministry of Environment and Food, DK-2600 Glostrup, Denmark

6: Department of Veterinary and Animal Sciences, University of Copenhagen, DK-1860 Frederiksberg C, Denmark

Short title: Evolution of SARS-CoV-2 in mink and humans in Denmark during 2020.

Corresponding authors: Thomas Bruun Rasmussen (tbru@ssi.dk) and Graham J. Belsham (grbe@sund.ku.dk)

Abstract

The severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2) has not only caused the COVID-19 pandemic but also had a major impact on farmed mink production in several European countries. In Denmark, the entire population of farmed mink (over 15 million animals) was culled in late 2020. During the period of June to November 2020, mink on 290 farms (out of about 1100 in the country) were shown to be infected with SARS-CoV-2. Genome sequencing identified changes in the virus within the mink and it is estimated that about 4000 people in Denmark became infected with these mink virus variants. Phylogenetic analysis revealed the generation of multiple clusters of the virus within the mink. A detailed analysis of the changes in the virus during replication in mink and, in parallel, in the human population in Denmark, during the same time period, has been performed here. The majority of cases in mink involved variants that had the Y435F substitution and the H69/V70 deletion within the Spike (S) protein; these changes emerged early on during the outbreak. However, further introductions of the virus, with variants lacking these changes, from the human population into mink also occurred. Based on phylogenetic analysis of the available viral genome data, we estimate that there were a minimum of about 17 separate examples of mink to human transmission of the virus in Denmark, using a conservative approach, but up to 60 such events (95% credible interval: (35-77)) were identified using parsimony to count cross-species jumps on transmission trees inferred using a Bayesian method. Using the latter approach, it was estimated that there were 136 jumps (95% credible interval: (112-164)) from

humans to mink. Thus, transmission of these viruses from humans to mink, mink to mink, from mink to humans and between humans were all observed. (296 words)

Author summary

In addition to causing a pandemic in the human population, SARS-CoV-2 also infected farmed mink. In Denmark, after the first identification of infection in mink during June 2020, a decision was made in November 2020 to cull all the farmed mink. Within this outbreak, mink on 290 farms (out of about 1100 in the country) were found to have been infected. We showed, by analysis of the viruses from the mink, that the viruses on the farms were mainly of three different, but closely related, types (termed Clusters 2, 3 and 4) that shared certain distinctive features. Thus, we found that many outbreaks in mink resulted from transmission of the virus between mink farms. However, we identified that new introductions of other virus variants, presumably from infected humans, also occurred. Furthermore, we showed that spread of the virus from infected mink to humans also happened on multiple occasions. Thus, transmission of these viruses from humans to mink, mink to mink, from mink to humans and between humans were all observed. (172 words)

Introduction

The severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2) has caused the COVID-19 pandemic [1], with over 675 million cases reported globally and it has contributed to the deaths of at least 6.8 million people [2]. The coronavirus (RaTG13), which has been found to be the most closely related to SARS-CoV-2, was detected in horseshoe bats (*Rhinolophus affinis)* in China [3], with about 1200 nucleotide (nt) differences between their

full-length RNA genomes of about 30,000 nt (ca. 96% identity). It is not known how the virus moved from these bats to humans or if there was an intermediate host [4, 5], as with civet cats for the SARS-CoV [6]. In addition to the effect of the continuing pandemic in humans, the same virus has also had a drastic impact on farmed mink production worldwide. Outbreaks of disease on mink farms, caused by infection with SARS-CoV-2, were initially identified, during April 2020, in the Netherlands (NL) [7]. These were followed closely (from June 2020) by outbreaks in Denmark (DK) [8], a country with one of the highest levels (about 40%) of global mink production, involving at that time over 1100 farms and a population of about 17 million mink [9]. Spread of SARS-CoV-2 into mink was also observed in a variety of other countries, including Canada, France, Greece, Italy, Lithuania, Spain, Sweden and the USA [10].

In total, SARS-CoV-2 infections were detected on 290 mink farm premises in DK (ca. 25% of the total) and this contributed to the Danish government's decision in early November 2020 to stop all mink production within DK [11]. The entire mink population was culled [9] and mink production halted until the end of 2022. The production of mink in the NL was also stopped in 2020, bringing forward an earlier planned end to this industry [12].

During the course of the outbreaks in mink in DK, a large number of different virus variants were observed. However, most of the viruses from mink that were analyzed had a specific mutation (A22920T) within the gene encoding the Spike (S) protein, resulting in the conservative amino acid substitution Y453F (tyrosine to phenylalanine), which occurred on the first mink farm found to have infected animals in DK [8]. This mutation was one of the defining changes that lead to the emergence of the virus pangolin lineage termed B.1.1.298 within the European Clade 20B. This same change was seen on one mink farm in the NL, early in the outbreak there, but also later in other mink farms [7, 13]. However, these variants, belonged to two different clades, 19A and 20A, and did not predominate in the NL.

The residue Y453 lies within the receptor binding domain (RBD) of the S protein that is known to interact with the cellular receptor, angiotensin-converting enzyme 2 (ACE2), which is used by the virus [14]. It has been reported that the Y453F substitution enhances binding of the virus to the mink ACE2 protein without compromising interaction with the human ACE2 protein [15].

A second, early, change in the viruses circulating in the mink population was the deletion of six contiguous nucleotides in the S gene coding sequence, which resulted in the loss of two amino acid residues, H69 and V70 (termed H69/V70del), from the S protein [16]. This change was first detected (in August 2020) on the 4th farm with infected mink in DK along with additional sequence changes, in other parts of the virus genome (including nucleotide changes leading to the amino acid substitutions P3395S in ORF1a and S2430I in ORF1b).

After the appearance of the Y453 and H69/V70del variants in mink, viruses with these changes were also found in the human population in the same region of DK, namely Northern Denmark [8, 11]. In total, the mink variants of SARS-CoV-2 were detected in over 1,100 people in DK out of 53,933 sequenced samples during the period from June 2020 to January 2021 [17] and this incidence was used to estimate that about 4000 humans in DK became infected with mink-derived viruses [11]. In Northern Denmark, where most SARS-CoV-2 outbreaks in mink occurred, amongst the people connected to mink farms, about 30% tested positive for SARS-CoV-2 in the period from June to November 2020 and approximately 27% of the SARS-CoV-2 samples from humans in this community were mink-associated [11].

During August and September 2020, mink on substantially more farms tested positive for SARS-CoV-2 [9]. This was coincident with extensive community spread of the virus [11] and further sequence changes generating multiple discrete clusters of viruses (termed Clusters

2, 3, 4 and 5) within the mink phylogeny (Figure 1). There was particular concern about a Cluster 5 isolate (named hCoV-19/Denmark/DCGC-3024/2020, GISAID EPI_ISL_616802), which had a number of amino acid sequence changes in the S protein (Y453F, I692V and M1229I as well as the H69/V70del). Preliminary testing of this virus isolate suggested a possible decrease in neutralization of this virus variant by human antibodies [18]. However, further analysis [19] showed that the impact of these changes on the ability of this virus to be neutralized by antibodies from convalescent humans was generally rather limited. Similarly, it has been found that there was very little loss of neutralization of pseudoviruses carrying a Cluster 5-like S protein, compared to wild-type, by sera from people twice vaccinated with Pfizer or Moderna mRNA vaccines [20].

In the current study, the genomic sequences of viruses from nearly all known infected mink farm premises in DK have been analyzed together with the sequences of the viruses circulating in the human population in DK during the same time period. This sheds light on the spread and evolution of the virus within mink and also describes many occasions when the virus was transmitted from humans to mink, as well as *vice-versa*.

Results

Appearance of multiple clusters of SARS-CoV-2 in mink

After the initial cases (starting in June 2020) of SARS-CoV-2 infection on four mink farms in DK [8, 16], there was further spread of the virus to other farms (Figure 1, Table 1). Outbreaks initially occurred within Northern Denmark but spread into Central and Southern Denmark (Figure 2). The virus variants found in mink in DK, during August and September 2020, all belonged to the same pangolin lineage, B.1.1.298, as for the initial cases, and were most likely descendants from the virus identified in the mink population in June. They all had the Y453F substitution in the S protein that was first observed on farm 1 [8]. It should be

noted that from farm 1 onwards, each farm with infected mink was numbered consecutively following detection of SARS-CoV-2 on the farm. The SARS-CoV-2 in DK at that time, in both humans and mink, all had the A23403G change (encoding the substitution D614G within the S protein) compared to the Wuhan strain and this change is not considered further.

Additional mutations emerged within the infected mink. Whole-genome-based phylogenetic analysis, using the maximum-likelihood method, performed on 698 sequences from infected mink (from nearly all the affected farms in DK), showed a segregation of the viruses from the initial cases into four major clusters (termed Clusters 2, 3, 4 and 5) indicating multiple transmission pathways (Figure 1 and Supplementary Figure 1). A circular representation of the phylogenetic tree clearly shows the general dominance of Clusters 2, 3 and 4 within this epidemic (Figure 1), but sequencing was only performed on a small subset of the infected mink, thus the precise proportions of mink infected with each variant is not known. A rectangular version of the phylogenetic tree based on the same set of virus sequences, but including sequence IDs and farm numbers, is shown in Supplementary Figure S1.

Viruses present on farms 1-4 [16], represent parental sequences to Clusters 2, 3, 4 and 5 (Supplementary Figure S1). In total, 270 of the 290 farms (i.e. 93%) that were tested positive for SARS-CoV-2 by the end of November had mink infected with variants of lineage B.1.1.298. Cluster 4 was the most common virus variant found amongst these outbreaks (Figure 1) and was detected on 121 farms, while Cluster 2 and Cluster 3 viruses were found on 76 and 66 farms, respectively (note, some farms had viruses from more than one cluster present, see Supplementary Figure S1). In contrast, the Cluster 5 variant was only observed in mink from five farms in Northern Denmark (Table 1 and Figure 2A) and only during the first part of September 2020, whereas the other Clusters persisted until the culling of all mink in

DK that ended in late November (Table 1). Further details of the various Clusters are described in Supplementary Information file S1.

The mink variant viruses with Y453F (within lineage B.1.1.298 including Clusters 2, 3, 4 and 5) clearly made up the majority of the variants found on Danish mink farms during the mink epidemic (Figure 1). However, new introductions of SARS-CoV-2 into mink also occurred, which lead to the C1-C8 variant groups. These new introductions occurred in multiple locations within Northern, Central and Southern Denmark (Figure 2B). These viruses are clearly distinct from the majority of those that infected the mink. For example, the viruses in C1-C8 lack the Y453F substitution in the S protein and they do not belong to the B.1.1.298 lineage. In total, mink on eighteen farms were infected with SARS-CoV-2 lineage variants other than B.1.1.298. These individual independent introductions are described in more detail in Supplementary Information file S2.

Evolution of SARS-CoV-2 in mink and humans

In order to investigate the evolution of SARS-CoV-2 in mink and in humans within DK, the sequences of the viruses from both hosts were compared. The full-genome sequences of SARS-CoV-2 from samples collected from Danish mink were collected from GISAID [21] and low-quality sequences (i.e. with more than 10 unresolved nucleotides) were removed. Sequences from humans in DK, circulating at the same time, were also retrieved. For each of the datasets, identical or nearly identical sequences were also removed (see Materials and Methods). The final data set comprised 258 sequences from mink on 129 farms and 497 sequences from humans across DK. These were aligned to the Wuhan-Hu-1 reference genome (GenBank accession no. NC_045512) as described, and a phylogenetic tree, including the mink and human viruses, was constructed (Figure 3). It is apparent that there was considerable heterogeneity among both the mink and human sequences in DK during this period. Furthermore, it can be seen that sequences derived from mink and human hosts are

interspersed on the tree, indicating multiple cross-species transmission events occurred (Figure 3).

Evolution and spread of mink-derived virus variants

At the time of the first introduction of SARS-CoV-2 into farm 1, in Northern Denmark (Figure 2A), the amino acid substitution Y453F, in the receptor-binding domain of the S protein (resulting from the mutation A22920T), had not been seen anywhere else (globally) except in mink from one of the infected mink farms in the NL. In this case, the substitution was in a different clade (19A) of SARS-CoV-2 [7, 8], so this finding did not indicate a connection between the outbreaks in DK and in the NL. Virus from the person connected to farm 1 in DK, who is presumed to be the source of the outbreak in mink, did not have this mutation in the spike protein gene. Indeed, the viruses from mink on farm 1 varied at this position, some had the A22920T mutation (resulting in the Y453F substitution) whereas others lacked this change [8] (Figure 1). Phylogenetic analysis based on whole-genome SARS-CoV-2 sequences from both mink and human hosts, also clearly showed that the Y453F substitution evolved only once (among mink on farm 1) and then spread, with all descendant mink- and human-derived sequences retaining this mutation (Figure 3 and 4).

The deletion of residues H69/V70 in the S protein, on the other hand, appears to have evolved up to 5 times independently among the human and mink viruses analyzed here (Figures 5 and 6). One of these events occurred among the group of viruses in the mink that already had the Y453F substitution. The H69/V70del modification, as well as two other deletions in ORF1a, were observed for the first time on farm 4 [16]. Specifically, and based on the clock-tree reconstructed using BEAST 2, the deletion resulting in the H69/V70del change evolved about 2-7 weeks after the appearance of the Y453F variant (Supplementary Figure S2). This is consistent with a previous analysis, which showed that deletion of H69/V70 from the S protein increases virus infectivity and compensates for an infectivity

defect resulting from the RBD-substitutions N439K and Y453F [22]. All viruses, in the clade descending from this event, inherited this deletion, which was, therefore, present in the vast majority of the mink-derived viruses analyzed here.

Among viruses, which do not have the Y453F substitution, the H69/V70 deletion appeared again in 4 separate locations on the phylogeny (Figures 5 and 6). Two of these are singleton human sequences, that are basal to the Danish sequences, and they may, therefore, represent separate introductions rather than cases where the deletion evolved among Danish viruses. In addition to these single leaves, there are two clades, within the non-Y453F part of the tree, where multiple related sequences all have the deletion (Figure 6). It appears that the deletion evolved independently among Danish viruses in these two cases, and then spread. One of these clades contains 3 human sequences, while the other contains 1 mink-sequence and 4 human sequences indicating that virus with the deletion was transferred between humans and mink. In some of these viruses, the H69/V70 deletion was coupled with the N439K substitution in the S protein, which is also within the RBD, and where the deletion has also been reported to function as a compensatory change [22].

Inference of the number of cross-species transmissions in DK

In previous studies, Wang et al. [23] defined criteria for identifying a cross-species transmission event for SARS-CoV-2 using a subset of Danish sequences. These criteria were: (1) that the direct two branches after the root of the clade have a different host; and (2) that the posterior probability of both branch and ancestral host for the root of the clade is >0.8. In the dataset used by Wang et al. [23], three independent cross-species transmission events were observed, all of which were caused by human-to-mink transmission. In addition, six SARS-CoV-2 sequences from humans were found to be very similar to mink-derived viral genomes, indicating they were most likely transmitted from mink to humans. However,

Wang et al. [23] could not determine, using their analyses, how many independent cross-species transmission events occurred due to the low posterior probabilities of the branches.

In order to further investigate the incidence of cross-species virus transmission events, the collected whole-genome sequences from DK (as described here) were used to infer the number of times that SARS-CoV-2 jumped from mink to humans (and *vice-versa*). Briefly, BEAST2 [24] was used to reconstruct clock model-based phylogenies. Then TransPhylo [25] was used to infer transmission trees based on the output from BEAST2, and finally the sumt and phylotreelib python packages [26,27] were used to analyze the transmission trees and count the likely number of zoonotic and reverse zoonotic jumps between the two species. This number was calculated using three different methods (see Materials & Methods). In method A, the number of inferred direct transmissions from an observed mink sequence to an observed human sequence were counted. Using this approach, it was estimated that there had been about 9 direct transmissions (posterior mean: 8.6; 95% credible interval: 6-11) from one of the 258 mink sequences included in the dataset, to one of the 497 human sequences. In method B, *indirect* transmissions were also inferred from an observed mink sequence, via an unobserved intermediate host, to an observed human sequence. Using this approach, it was estimated that there had been about 17 jumps (posterior mean: 17.3, 95% credible interval: 14-21) from one of the mink to one of the humans in the data set. Using this same method, there were estimated to be about 18 jumps (posterior mean: 18.3; 95% credible interval: 14-21) from humans to mink. Finally, in method C, the number of cross-species jumps was estimated using a parsimony method applied to the TransPhylo output, including inferred unobserved mink and human hosts also. Using this approach, it was found that there had been about 60 jumps from mink to humans in DK during the investigated period (posterior mean: 59.6; 95% credible interval: 35-77). The result of method B, about 17 jumps from mink to humans, can be considered as a fairly high-confidence, but conservative, estimate, i.e., it is

reasonably sure that the number of jumps is not less than this. However, since the virus from only a small proportion of the infected mink that were in DK during that time have been sequenced, it is almost certain that many interspecies jumps will be missed. The result from method C, i.e. about 60 jumps, may be argued to be probably closer to the real number as it represents a less conservative estimate. However, it comes with a greater uncertainty.

Using method C, a parsimony method applied to the TransPhylo output, it was estimated that there had also been about 136 jumps from humans to mink (posterior mean: 135.5, 95% credible interval: 112-164). This fits fairly well with the 129 different mink farms, with infected mink, represented in our data set, since it is assumed that most of the virus introductions into the mink farms have occurred by independent human-to-mink transmission events (not by mink from one farm directly infecting mink at another farm).

Discussion

SARS-CoV-2 infection of farmed mink in DK contributed to the epidemic in humans in DK during 2020. The epidemic in mink was not being efficiently controlled by the measures taken (mink on 290 farms out of about 1100 in the country were found to have been infected) and it was decided to cull over 15 million mink. This resulted in the closure of the mink production industry until after the end of 2022. Most of the outbreaks in mink were caused by one of three different virus lineages, termed Clusters 2, 3 and 4, all of which belong to the pangolin lineage B.1.1.298 (Figure 1). These clusters shared some common features, namely the H69/V70del and Y453F changes, within the S protein. The deletion of H69/V70 has arisen independently in a variety of different lineages of SARS-CoV-2, both within mink and human variants. The deletion is associated with increased cleavage of the S protein and confers enhanced virus infectivity [22].

A virus isolate from Cluster 5, with additional amino acid changes, was the focus of considerable attention since preliminary studies indicated this isolate showed resistance to neutralization by antibodies from a small panel of convalescent human patients [18]. However, in follow up studies [19], it was found that the antibodies from just 3 out of 44 patient samples tested had a >3-fold reduction in virus neutralization titer against the Cluster 5 virus isolate compared to a virus from early in the pandemic. Only one sample from the 44 patients had a neutralization titer that was reduced by 4-fold or more [19]; the latter being the threshold set for defining neutralization resistance [28].

The Y453F substitution was found to have evolved only once in the mink in DK, on farm 1 [8]. This change was present in the majority of the sampled mink sequences (Figure 1 and Suppl. Figure S1) and was also found in sequences from more than 1100 human cases in DK. It has been estimated that about 4000 humans have been infected with this variant [11]. Thus the Y453F change clearly does not have a severely detrimental effect on the ability of the virus to infect humans [29]. However, viruses with this change were rapidly lost following the culling of all the mink (Table 1 and [11]). Cluster 5 viruses were not detected in mink or humans after mid-Sept. 2020 but viruses of the B.1.1.298 lineage (with the Y453F change) were detected in humans until January 2021 [17]. This suggests that viruses with the Y453F change did not have a selective advantage in humans at this time point. However, the generation of the Y453F variant (with the H69/V70del) in a patient with lymphoma has been reported [30], in a virus lineage separate from the mink viruses. As indicated above, the Y453F change only occurred once in mink in DK, on farm 1 [8], and was then retained in all descendant viruses analyzed here. However, it is notable that this change also has occurred independently in other mink virus sequences in the NL [7], Poland [31], the USA [32] and (based on sequences from GISAID [21]) in Lithuania and Latvia. All of these changes occurred in lineages other than B.1.1.298, indicating convergent evolution due to selective

advantages in mink. It should be noted that all but one sequence within the B.1.1.298 lineage originated from DK [21]. The single sequence from outside DK was found in a human sample collected in the Faroe Islands in September 2020.

In the lineage C4, which was first recognized in mid-October 2020 (i.e. shortly before the cull commenced) and lacks the Y453F change, another change, N501T, was detected on multiple farms (Supplementary Information file S2). Like the Y453F change, this substitution occurs at the interface between the ACE2 receptor and the S protein. Thus, it may achieve a similar effect [29]. It is notable that this change has also occurred in mink sequences from multiple countries and in different virus lineages as for the Y453F substitution (see above).

It is most likely that the initial introductions of SARS-CoV-2 into mink farms occurred from infected people. It is apparent that the virus, having acquired the Y453F change, then spread quickly and easily within the mink [8, 16]. Transmission from mink back into the human population clearly occurred too.

Assessing the extent of interspecies virus transmission is not simple, see Wang et al. [23]. Due to the many highly similar sequences, there will be several branches in the phylogenetic tree with poor support, and this causes what may be termed an entropic problem leading to an upward bias in the count of interspecies jumps [33]. If a set of, say, 5 mink sequences and 5 human sequences each have one unique mutation, then their pairwise distances will all be 2, and all the possible resolutions of this 10-leaf subtree will be equally likely. However, since there are many more possible subtrees where the 5 mink and 5 human leaves are intermingled, than there are possible subtrees where they are cleanly separated, then the average number of inferred jumps will be biased towards more than 1 inter-species jump, even though the data would also be consistent with only one zoonotic event. This means that ordinarily used methods for dealing with phylogenetic uncertainty, such as performing the computation on all or many trees from BEAST's posterior sample, will not

work (instead of getting a reliable posterior count, accounting for the uncertainty, the inclusion of less supported trees will create a bias for over-counting).

Here, we have used three different methods to assess interspecies virus transmission. Using method B, the analysis of the sequences indicated that at least 17 (95% credible interval: 14-21) different mink to human transmission events have happened in DK. This was estimated using a very conservative approach. Using an alternative method, based on analyzing the output from TransPhylo using parsimony (termed here method C), about 60 jumps from mink to humans were estimated to have occurred. Furthermore, this methodology generated an estimate of 135 jumps from humans to mink. This number fits well with the 129 farms represented in the data set that had infected mink. The transmission of the mink variant viruses from one mink farm to another occurred very efficiently. However, the mechanisms involved in this spread are not established [9]. In many cases, it may have been by human contacts with multiple mink farms but other routes are also possible. It is assumed that most of the introductions of the virus onto these mink farms have occurred by independent mink-to-human and then human-to-mink transmission events (not by mink from one farm infecting mink at another farm). Airborne transmission of the virus from mink farms to humans not connected to the farm seems unlikely, since the concentration of virus in the air outside of the mink farms appears to be low [9]. However, this topic deserves further study. The major proportion of the viruses that infected mink in DK had the Y453F substitution together with the H69/V70del in the S protein, including all of the viruses in Clusters 2, 3, 4 and 5 (Figure 1). This suggests that, although new introductions of the virus from humans occurred (as with C1-C8), these were much less important for the total outbreak in mink than the mink farm to mink farm transmission. However, it is clearly not possible to know whether some of these virus variants would have become predominant among the mink if they had not been culled.

Concluding remarks

It is apparent that SARS-CoV-2 readily infected farmed mink and spread quickly between farms. Transmission from infected humans to mink and from infected mink to humans occurred on multiple occasions and the mink-derived viruses then spread among people. There were legitimate concerns that replication of SARS-CoV-2 in a large population of mink could generate novel variants that would have adverse effects on human health due to antigenic change, greater transmissibility or higher fitness. However, mink-derived viruses with such unwelcome characteristics did not spread among humans before the mink population was culled. Variants of SARS-CoV-2 that did arise in mink (e.g. with the changes Y453F and H69/V70del in the S protein) were transmitted to, and within, the human population but died out either before, or soon after, the culling of the mink population in DK.

Materials and Methods

Sequencing strategy

Whole genome amplification of SARS-CoV-2 in mink and human samples was performed using a modified ARTIC tiled PCR protocol (see [34]) with amplicons ranging from 1000-1500 bp. A custom 2-step PCR with barcoding was applied to the amplicon libraries, then the libraries were normalized, pooled, and sequenced using Oxford Nanopore's SQK-LSK109 ligation kit on a MinION device with R.9.4.1 flowcells. The full protocol is available [35].

Construction of maximum likelihood phylogenetic tree

The maximum likelihood phylogeny of all 698 SARS-CoV-2 sequences from mink isolates was reconstructed using IQ-TREE version 2.0.3 [36] with a GTR model, based on the

alignment obtained by comparing each sequence to the Wuhan-Hu-1 reference genome (GenBank accession no. NC_045512) using MAFFT version 7.475 [37] with option '--addfragments'. The phylogenetic tree was thereafter annotated using package ggtree in R version 4.2.1 [38]. Clusters 2-5 were derived from the initial cases (on farms 1-5) while the separate introductions that resulted in the C1-C8 variant groups were defined from a phylogeny based on human and mink sequences by picking the smallest possible monophyletic group containing one or more mink sequences.

Construction of Bayesian phylogenetic trees

Whole-genome sequences of SARS-CoV-2 derived from infected farmed mink and humans in DK were collected from GISAID [21] on August 31st 2023. Sequences derived from mink were collected by searching for complete sequences passing GISAID's high coverage filter (allowing only entries with <1% Ns and <0.05% unique AA mutations) with a precise collection date. These gave rise to dataset 1; for this dataset, consisting of mink virus sequences, duplicate sequences derived from samples from the same farm on the same date were removed. Similarly, sequences derived from humans were collected by searching for complete sequences with a collection date between June 1st 2020 and February 28th 2021 passing GISAID's high coverage filter. Two different datasets were constructed consisting of human virus sequences: dataset 2 with the amino acid substitution S:Y453F and dataset 3 without the amino acid substitution S:Y453F. For datasets 2 and 3, duplicate sequences were removed if they were sampled on the same day. This was done to preserve the temporal signal in the data.

Sequences with more than 10 undetermined nucleotides were removed from the datasets, and the datasets were pre-processed by masking as described [39], removing sequences with more than 100 end gaps. Dataset 3 was further reduced to minimize the computational load using CD-HIT-EST from CD-HIT [40] to achieve a representative dataset

using a similarity threshold of 0.999. The three datasets were combined into one consisting of 258 sequences from mink (derived from 129 mink farms), 49 sequences from humans without the S:Y453F substitution and 448 sequences from humans with the S:Y453F substitution. These sequences were aligned as described above.

Estimation of the number of zoonotic jumps from mink to human

To determine transmission pathways, information from the phylogenies together with the relative sampling dates was combined. Phylogenetic trees were reconstructed using BEAST 2 [24]. The substitution model was GTR with empirical base frequencies and gamma-distributed rates with 4 discrete categories, combined with a strict molecular clock model calibrated by using the sequence sampling-dates, obtained from GISAID, to date the tips of the tree. The tree prior was the birth-death skyline serial model, with 10 dimensions for the reproductive number parameter, and one dimension for the sampling proportion [41]. The model estimates a separate effective reproduction number for each of 10 equally large time-intervals covering the time-span from the root of the tree to the farthest tip. The prior for the becoming-uninfectious rate parameter was lognormal(M=52.0, S=1.25, mean in real space) per year, corresponding to a prior 95% credible interval of [1.3, 180] days for the duration of an infectious period. The prior for the clockrate was lognormal(M=0.001, S=1.25, mean in real space) substitutions per site per year, corresponding to a 95% prior interval of [4.0E-5, 5.3E-3] substitutions per site per year. Both of these priors are weakly informative and help to regularize model fitting without imposing very strict constraints on the estimated values for these parameters. Other priors were left at their default values. Two parallel MCMC chains were run for 50 million iterations each with logging of trees and other parameters every 4000 iterations (for a total of 2 x 12,500 parameter samples). A burn-in of 30% (15 million generations) was used. The software Tracer v1.7.2 [42] was used to analyze

parameter samples. Marginal posterior distributions from the two runs were essentially identical, indicating good convergence. Effective sample sizes for all parameters were well above 200, except for the following: posterior (ESS=166), likelihood (ESS=94), tree-length (ESS=136), BDSKY_serial (ESS=138). The software phylotreelib [26] and sumt [27] were used to analyze tree-samples, and to extract post-burnin trees and compute maximum clade credibility trees. Tree samples from the two independent runs were very similar, with average standard deviation of split frequencies (ASDSF) of 0.0125. The number of effective tree samples was estimated by first computing the log clade credibility for each tree-sample (based on clade frequencies from all post-burnin trees), and then using Tracer to compute ESS from this proxy measure [43]. Computed this way, the tree-sample ESS was 287, indicating an acceptable number of independent tree samples in the posterior.

To infer transmission trees, the software TransPhylo v1.4.10 [25] was used. This takes as input a pre-computed, dated phylogeny, where leaves correspond to pathogens sampled from the known infected hosts. The main output is a transmission tree that indicates "who" infected "whom", including the potential existence of unsampled individuals who may have acted as missing transmission intermediates. For input we used the maximum clade credibility (MCC) tree with common-ancestor depths. A further 28 other trees from BEAST2's posterior samples were analyzed, chosen to cover a range of different log-clade credibility values. We also used common-ancestor depths to set the branch lengths of these trees. Before analyzing any of these trees, the original Wuhan sequence was removed from the tree with the aim of having a more homogeneous substitution process on the remaining branches for the TransPhylo analysis. The generation time distribution in TransPhylo was set to be gamma-distributed with shape-parameter=60 and scale-parameter=0.0004105. These parameters were chosen to match the posterior 95% credible interval, found in the BEAST-analysis, as closely as possible (6.86 to 11.4 days). The parameters were found using the

optimize.minimize function from the SciPy python package [44]. TransPhylo was run for 10 million iterations, sampling every 2000 generations, and using a burnin of 50%. This gave a total of 2500 post-burnin samples of transmission trees and other parameters, for the MCC tree and for each of the 28 other trees from the BEAST posterior sample. For the TransPhylo run, we set updateOff.p=TRUE to allow estimation of the offspring distribution. Convergence was checked by inspecting trace plots and computing ESS.

The output from TransPhylo was further analyzed to estimate the number of times SARS-CoV-2 jumped between mink and humans. This was done by inspecting each of the 72,500 posterior transmission-tree samples (i.e. 29 times 2500), and for each of them counting the number of jumps in three different ways. In method A: the inferred direct transmissions from an observed mink sequence to an observed human sequence were counted (i.e., cases where TransPhylo inferred that both the source and the target of a cross-species transmission event were included in the data set). In method B: the number of inferred *indirect* transmissions from an observed mink sequence to an observed human sequence were counted. Occasionally TransPhylo will infer transmission chains that include one or more unobserved links (e.g., mink -> unknown -> unknown -> human), and these, of course, also imply transmission of the virus from a mink to a human somewhere in that chain. In method C: a parsimony method was used to infer the minimum number of mink-to-human transmissions based on the posterior sample of the transmission trees inferred by TransPhylo. Specifically, the algorithm of Hartigan [45] was implemented in a version that allowed some internal nodes on the tree to be observed (i.e., their state sets are simply taken to be the observed host for that internal node).

Pangolin lineage determination

The Pangolin lineage for the individual variants has been determined by analysis of the mink sequences in the database PANGO lineages [46]. In addition, a search in the GISAID EpiCoV database [21] has been used for further analysis in order to examine the occurrence of selected variants in published mink sequences and human sequences. When describing the observed changes in the S protein, the change D614G (compared to the reference Wuhan strain) was omitted, as this change occurred very early in the pandemic and is present in all sequences during the period of interest.

Author contributions

Conceptualization: T.B.R., A.Bø., G.J.B.; Data curation: T.B.R., A.S.O.; Formal analysis: T.B.R., A.G.Q, A.G.P., M.J.F.G., J.F., M.R.; Funding acquisition: A.Bø., T.B.R., G.J.B., A.G.P.; Investigation: T.B.R., R.N.S., M.S., L.L. A.Bo.; Methodology: R.N.S., M.S., A.G.Q., A.G.P.; Resources: L.L., S.M. (study materials); Software: R.N.S., A.G.P., A.G,Q, ; Supervision: T.B.R., A.G.P., A.Bø.; Validation: T.B.R.; Visualization: T.B.R., A.G.Q., F.F.C.A., M.J.F.G., E.R.T., R.N.S., A.G.P., A.Bo.; Writing- original draft preparation: G.J.B; Writing- review & editing: All authors.

Acknowledgements

We gratefully acknowledge all data contributors, i.e., the Authors and their Originating laboratories responsible for obtaining the specimens, and their Submitting laboratories for generating the genetic sequence and metadata and sharing via the GISAID Initiative, on which this research is partially based. We thank Kåre Mølbak for helpful comments on an early draft of the manuscript.

This research was funded by the Danish Veterinary and Food Administration (FVST) as part of the agreement for commissioned work between the Danish Ministry of Food and

519 Agriculture and Fisheries and the University of Copenhagen and the Statens Serum Institut.
520 Further funding has been received from the Danish National Research Foundation (grant
521 number DNRF170 to A.G.P).

522

523

References

1. World Health Organization. https://www.who.int/emergencies/diseases/novel-coronavirus-2019

2. Johns Hopkins University. Coronavirus resource center https://coronavirus.jhu.edu/ [accessed 2022 October 21]

3. Zhou P, Yang XL, Wang XG, Hu B, Zhang L, Zhang W, et al.. A pneumonia outbreak associated with a new coronavirus of probable bat origin. Nature. 2020;579: 270-273. doi: 10.1038/s41586-020-2012-7.

4. Boni MF, Lemey P, Jiang X, Lam TT, Perry BW, Castoe TA, et al. Evolutionary origins of the SARS-CoV-2 sarbecovirus lineage responsible for the COVID-19 pandemic. Nat Microbiol. 2020;5 :1408-1417. doi: 10.1038/s41564-020-0771-4.

5. Pekar JE, Magee A, Parker E, Moshiri N, Izhikevich K, Havens JL, et al. The molecular epidemiology of multiple zoonotic origins of SARS-CoV-2. Science. 2022;377: 960-966. doi: 10.1126/science.abp8337.

6. Wang LF, Eaton BT. Bats, civets and the emergence of SARS. Curr Top Microbiol Immunol. 2007;315: 325-44. doi: 10.1007/978-3-540-70962-6_13.

7. Oreshkova N, Molenaar RJ, Vreman S, Harders F, Oude Munnink BB, Hakze-van der Honing RW, et al. SARS-CoV-2 infection in farmed minks, the Netherlands, April and May 2020. Euro Surveill. 2020;25: 2001005. doi: 10.2807/1560-7917.ES.2020.25.23.2001005

544 **8.** Hammer AS, Quaade ML, Rasmussen TB, Fonager J, Rasmussen M, Mundbjerg K, et
545 al. SARS-CoV-2 Transmission between Mink (Neovison vison) and Humans,
546 Denmark. Emerg Infect Dis. 2021; 27: 547-551. doi: 10.3201/eid2702.203794.

547 **9.** Boklund A, Hammer AS, Quaade ML, Rasmussen TB, Lohse L, Strandbygaard B, et
548 al. SARS-CoV-2 in Danish Mink Farms: Course of the Epidemic and a Descriptive
549 Analysis of the Outbreaks in 2020. Animals (Basel). 2021;11:164. doi:
550 10.3390/ani11010164.

551 **10.** SARS-CoV-2 in animals used for fur farming. 2021.
552 https://www.woah.org/app/uploads/2021/03/glews-risk-assessment-fur-animals-sars-
553 cov-2.pdf

554 **11.** Larsen HD, Fonager J, Lomholt FK, Dalby T, Benedetti G, Kristensen B, et al.
555 Preliminary report of an outbreak of SARS-CoV-2 in mink and mink farmers
556 associated with community spread, Denmark, June to November 2020. Euro Surveill.
557 2021;26: 2100009. doi: 10.2807/1560-7917.ES.2021.26.5.210009.

558 **12.** https://promedmail.org/promed-post/?id=7994061.

559 **13.** Oude Munnink BB, Sikkema RS, Nieuwenhuijse DF, Molenaar RJ, Munger E,
560 Molenkamp R, et al. Transmission of SARS-CoV-2 on mink farms between humans
561 and mink and back to humans. Science. 2021;371: 172-177. doi:
562 10.1126/science.abe5901.

563 **14.** Yan R, Zhang Y, Li Y, Xia L, Guo Y, Zhou Q. Structural basis for the recognition of
564 SARS-CoV-2 by full-length human ACE2. Science. 2020;367: 1444-1448. doi:
565 10.1126/science.abb2762.

566 **15.** Ren W, Lan J, Ju X, Gong M, Long Q, Zhu Z, et al. Mutation Y453F in the spike
567 protein of SARS-CoV-2 enhances interaction with the mink ACE2 receptor for host
568 adaption. PLoS Pathog. 2021;17: e1010053. doi: 10.1371/journal.ppat.1010053.
569 **16.** Rasmussen TB, Fonager J, Jørgensen CS, Lassaunière R, Hammer AS, Quaade ML, et
570 al. Infection, recovery and re-infection of farmed mink with SARS-CoV-2. PLoS
571 Pathog. 2021;17: e1010068. doi: 10.1371/journal.ppat.1010068.
572 **17.** https://www.covid19genomics.dk/statistics
573 **18.** Lassaunière R, Fonager J, Rasmussen M, Frische A, Polacek Strandh C, Rasmussen
574 TB, et al. SARS-CoV-2 Spike Mutations Arising in Danish Mink and Their Spread to
575 Humans. København: Statens Serum Institut. 2020. . https://files.ssi.dk/Mink-cluster-
576 5-short-report_AFO2
577 **19.** Lassaunière R, Fonager J, Rasmussen M, Frische A, Polacek C, Rasmussen TB, et
578 al. *In vitro* Characterization of Fitness and Convalescent Antibody Neutralization of
579 SARS-CoV-2 Cluster 5 Variant Emerging in Mink at Danish Farms. Front Microbiol.
580 2021;12: 698944. doi: 10.3389/fmicb.2021.698944.
581 **20.** Garcia-Beltran WF, Lam EC, St Denis K, Nitido AD, Garcia ZH, Hauser BM, et al.
582 Multiple SARS-CoV-2 variants escape neutralization by vaccine-induced humoral
583 immunity. Cell. 2021;184: 2372-2383.e9. doi: 10.1016/j.cell.2021.03.013.
584 **21**. Shu, Y. and McCauley, J. GISAID: from vision to reality. Euro Surveill. 2017; 22:13
585 doi: 10.2807/1560-7917.ES.2017.22.13.30494
586 **22.** Meng B, Kemp SA, Papa G, Datir R, Ferreira IATM, Marelli S, et al. Recurrent
587 emergence of SARS-CoV-2 spike deletion H69/V70 and its role in the Alpha variant
588 B.1.1.7. Cell Rep. 2021;35: 109292. doi: 10.1016/j.celrep.2021.109292.
589

590 **23.** Wang L, Didelot X, Bi Y, Gao GF. Assessing the extent of community spread caused
591 by mink-derived SARS-CoV-2 variants. Innovation (Camb). 2021; 2(3):100128. doi:
592 10.1016/j.xinn.2021.100128.
593 **24.** Bouckaert R., Heled J., Kühnert D., Vaughan T., Wu C-H., Xie D., Suchard MA.,
594 Rambaut A., Drummond A. J. BEAST 2: A Software Platform for Bayesian
595 Evolutionary Analysis. PLoS Computational Biology. 2014; 10(4): e1003537.
596 doi:10.1371/journal.pcbi.1003537
597 **25.** Didelot X, Fraser C, Gardy J, Colijn C. Genomic Infectious Disease Epidemiology in
598 Partially Sampled and Ongoing Outbreaks. Mol Biol Evol. 2017; 34(4):997-1007. doi:
599 10.1093/molbev/msw275.
600 **26.** Pedersen A.G. phylotreelib: a python library for analyzing and manipulating
601 phylogenetic trees (Version 1.27.0) [Computer software]. 2023.
602 https://doi.org/10.5281/zenodo.10148565
603 **27.** Pedersen, A.G. sumt: a command-line program for computing consensus trees and
604 other phylogenetic tree summaries (Version 3.8.1) [Computer software]. 2023.
605 https://doi.org/10.5281/zenodo.10148693
606 **28.** Li Q, Wu J, Nie J, Zhang L, Hao H, Liu S, et al. The Impact of Mutations in SARS-
607 CoV-2 Spike on Viral Infectivity and Antigenicity. Cell. 2020;182: 1284-1294.e9.
608 doi: 10.1016/j.cell.2020.07.012.
609 **29.** Zhou J, Peacock TP, Brown JC, Goldhill DH, Elrefaey AME, Penrice-Randal R. et al.
610 Mutations that adapt SARS-CoV-2 to mink or ferret do not increase fitness in the
611 human airway. Cell Rep. 2022;38:110344. doi: 10.1016/j.celrep.2022.110344.
612 **30.** Stanevich OV, Alekseeva EI, Sergeeva M, Fadeev AV, Komissarova KS, Ivanova AA
613 et al. SARS-CoV-2 escape from cytotoxic T cells during long-term COVID-19. Nat
614 Commun. 2023;14(1):149. doi: 10.1038/s41467-022-34033-x.

615 **31.** Domańska-Blicharz K, Orłowska A, Smreczak M, Niemczuk K, et al. Mink SARS-
616 CoV-2 Infection in Poland - Short Communication. J Vet Res. 2021;65: 1-5. doi:
617 10.2478/jvetres-2021-0017.
618 **32.** Cossaboom CM, Wendling NM, Lewis NM, Rettler H, Harvey RR, et al. One Health
619 Investigation of SARS-CoV-2 in People and Animals on Multiple Mink Farms in
620 Utah. Viruses. 2022;15: 96. doi: 10.3390/v15010096.
621 **33**. Duchêne S, Lanfear R. Phylogenetic uncertainty can bias the number of evolutionary
622 transitions estimated from ancestral state reconstruction methods. J Exp Zool B Mol
623 Dev Evol. 2015;324(6): 517-524. doi: 10.1002/jez.b.22638.
624 **34.** Quick J. nCoV-2019 sequencing protocol V.1. 2020.
625 dx.doi.org/10.17504/protocols.io.bbmuik6w
626 **35.** Sørensen EA, Karst SM, Knutsson S. AAU-nCoV-2019 tailed long amplicon
627 sequencing. protocols.io; 2020. .dx.doi.org/10.17504/protocols.io.bfc3jiyn
628 **36.** Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: A Fast and
629 Effective Stochastic Algorithm for Estimating Maximum-Likelihood
630 Phylogenies, Molecular Biology and Evolution.2015;32:268–
631 274. https://doi.org/10.1093/molbev/msu300
632 **37.** Katoh K, Standley DM. MAFFT Multiple Sequence Alignment Software Version 7:
633 Improvements in Performance and Usability, Molecular Biology and
634 Evolution.2013;30: 772–780. https://doi.org/10.1093/molbev/mst010
635 **38.** Yu G, Smith DK, Zhu H, Guan Y, Lam TT.-Y. ggtree: an r package for visualization
636 and annotation of phylogenetic trees with their covariates and other associated data.
637 Methods Ecol Evol. 2017; 8: 28-36. https://doi.org/10.1111/2041-210X.12628

638 **39.** De Maio N, Walker C, Borges R, Weilguny L, Slodkowicz G et al. Masking strategies
639 for SARS-CoV-2 alignments, 2020. https://virological.org/t/masking-strategies-for-
640 sars-cov-2-alignments/480
641 **40**. Li W, Godzik A. Cd-hit: a fast program for clustering and comparing large sets of
642 protein or nucleotide sequences. Bioinformatics. 2006; 22: 1658–1659.
643 **41.** Stadler T, Kühnert D, Bonhoeffer S, Drummond AJ. Birth-death skyline plot reveals
644 temporal changes of epidemic spread in HIV and hepatitis C virus (HCV). Proc Natl
645 Acad Sci USA. 2013;110(1):228-33. doi: 10.1073/pnas.1207965110.
646 **42.** Rambaut A, Drummond AJ, Xie D, Baele G and Suchard MA. Posterior
647 summarisation in Bayesian phylogenetics using Tracer 1.7. *Systematic Biology*. 2018;
648 syy032. doi:10.1093/sysbio/syy032
649 **43.** Lanfear R, Hua X, Warren DL. Estimating the Effective Sample Size of Tree
650 Topologies from Bayesian Phylogenetic Analyses. Genome Biol Evol. 2016 Aug
651 16;8(8):2319-32. doi: 10.1093/gbe/evw171.
652 **44.** Virtanen P, Gommers R, Oliphant TE, Haberland M, Reddy T, Cournapeau D, et al.
653 SciPy 1.0 Contributors. SciPy 1.0: fundamental algorithms for scientific computing in
654 Python. Nat Methods. 2020; 17(3): 261-272. doi: 10.1038/s41592-019-0686-2.
655 **45.** Hartigan, JA. Minimum mutation fits to a given tree. Biometrics 1973; 29, 53.
656 **46.** Rambaut A, Holmes EC, O'Toole Á, Hill V, McCrone JT, Ruis C, et al. A dynamic
657 nomenclature proposal for SARS-CoV-2 lineages to assist genomic
658 epidemiology. Nat Microbiol. 2020; 5: 1403–1407. https://doi.org/10.1038/s41564-
659 020-0770-5
660
661

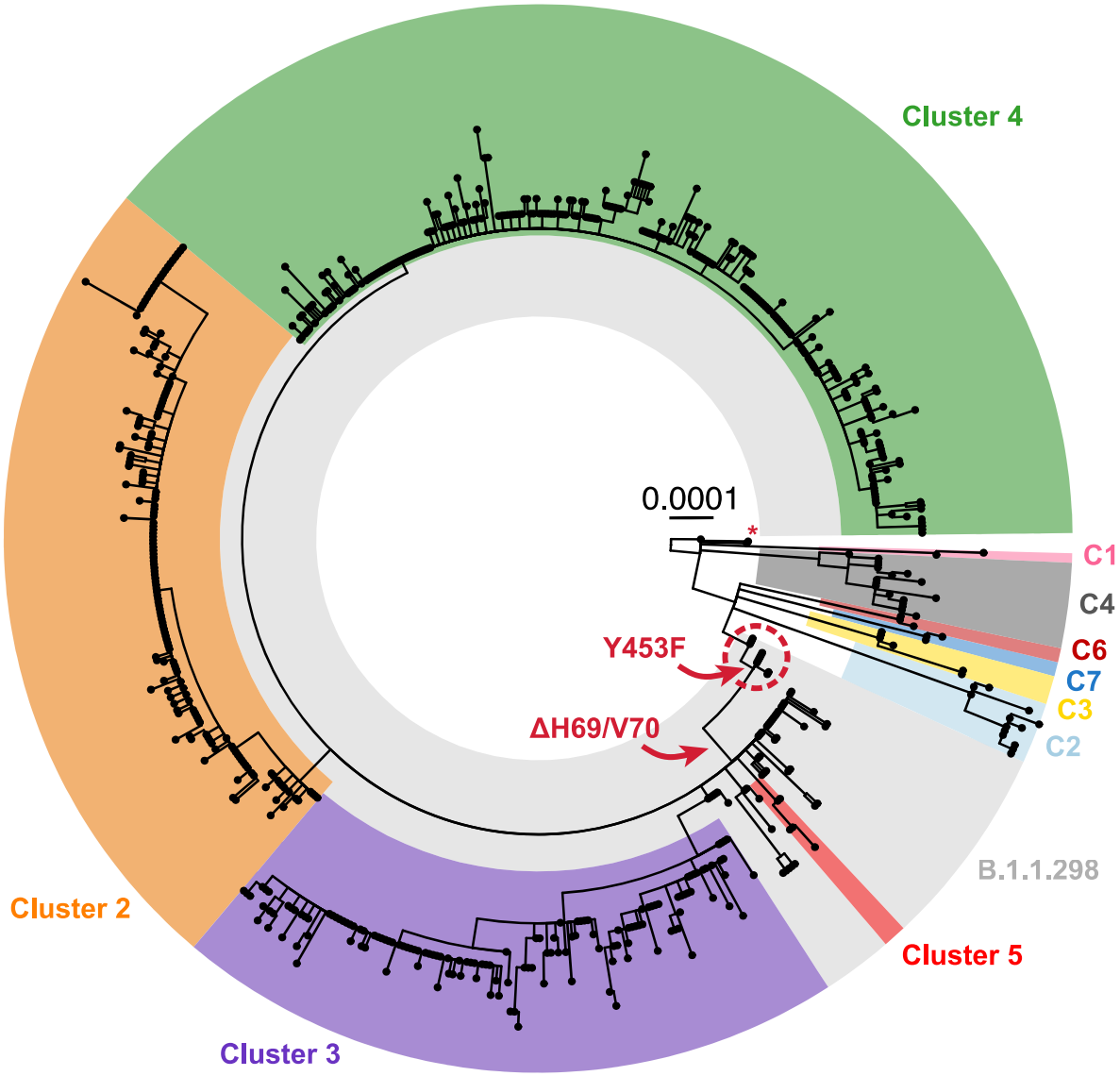
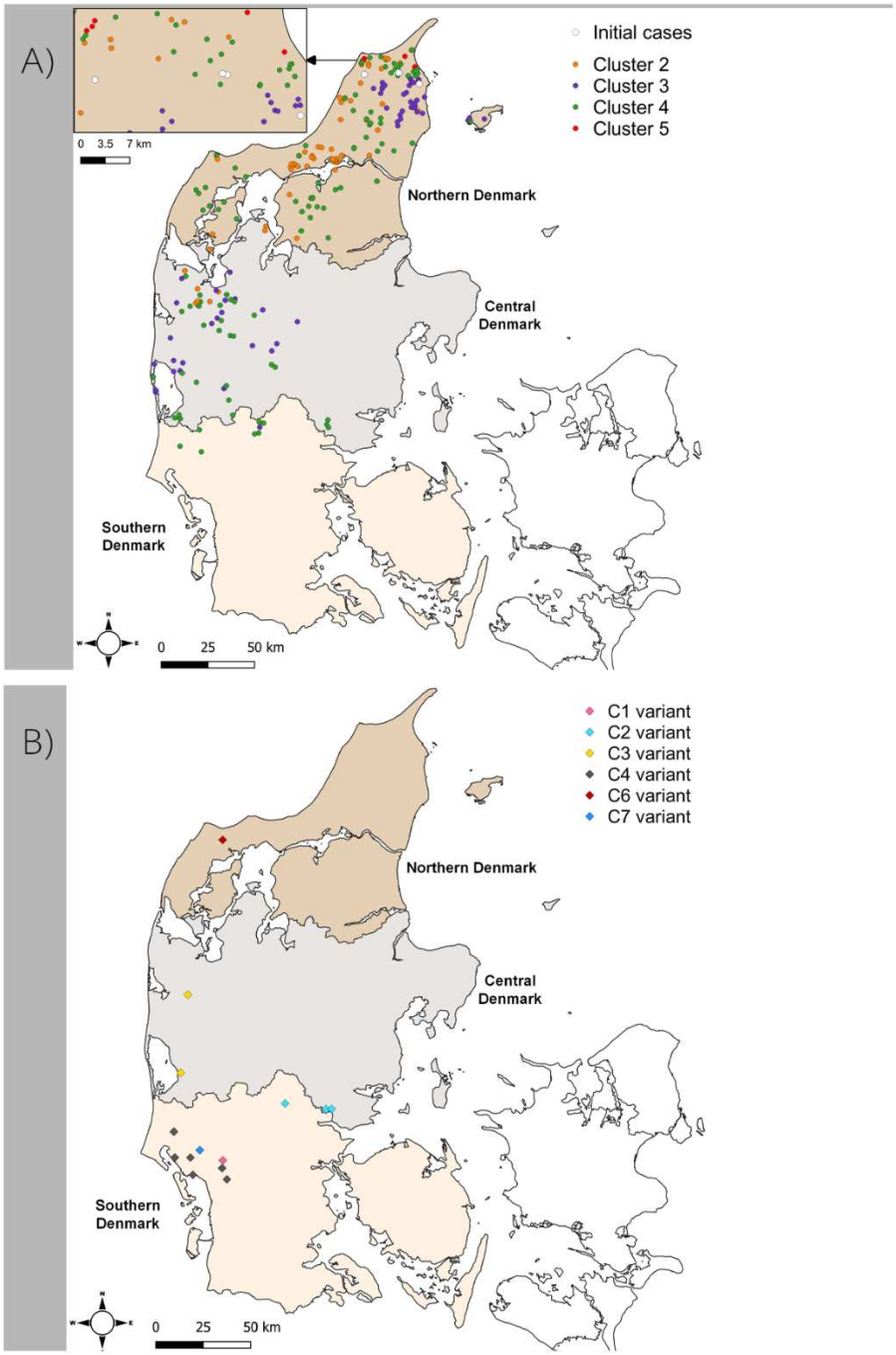


Figure 1. Phylogeny of the 698 SARS-CoV-2 whole-genome sequences from Danish mink. The majority of viruses found on infected farms, including those from the initial cases (farms 1-3, indicated within a red dashed circle) and viruses in Clusters 2-5, belong to pangolin lineage B.1.1.298 and are highlighted in light grey. Clusters 2-5 and viruses subsequently found as further spillovers from humans (C1-C4 and C6-C7) are highlighted in different colours. A singleton sequence belonging to C8 is indicated by a red asterisk. The occurrence of key sequence changes that were present in most mink virus sequences are indicated with red arrows. The scale bar indicates number of substitutions per variable site. The phylogeny was rooted with the basal reference sequence (NC_45512.1/EPI_ISL_406798, known as the Wuhan-Hu-1 virus) as the outgroup.



677

678 **Figure 2. Location of different SARS-CoV-2 variants in mink during the epidemic in**
679 **Denmark, June-November 2020.** Panel A. The location of the initial cases of SARS-CoV-2
680 infection in Northern Denmark are indicated. Subsequently, further cases occurred and the
681 virus diverged, within lineage B.1.1.298, into Clusters 2, 3, 4 and 5 (as shown in Figure 1).
682 Panel B. Later in the epidemic, new introductions of viruses from different lineages occurred
683 and these are named as C1-C7 (see Table 1).

684

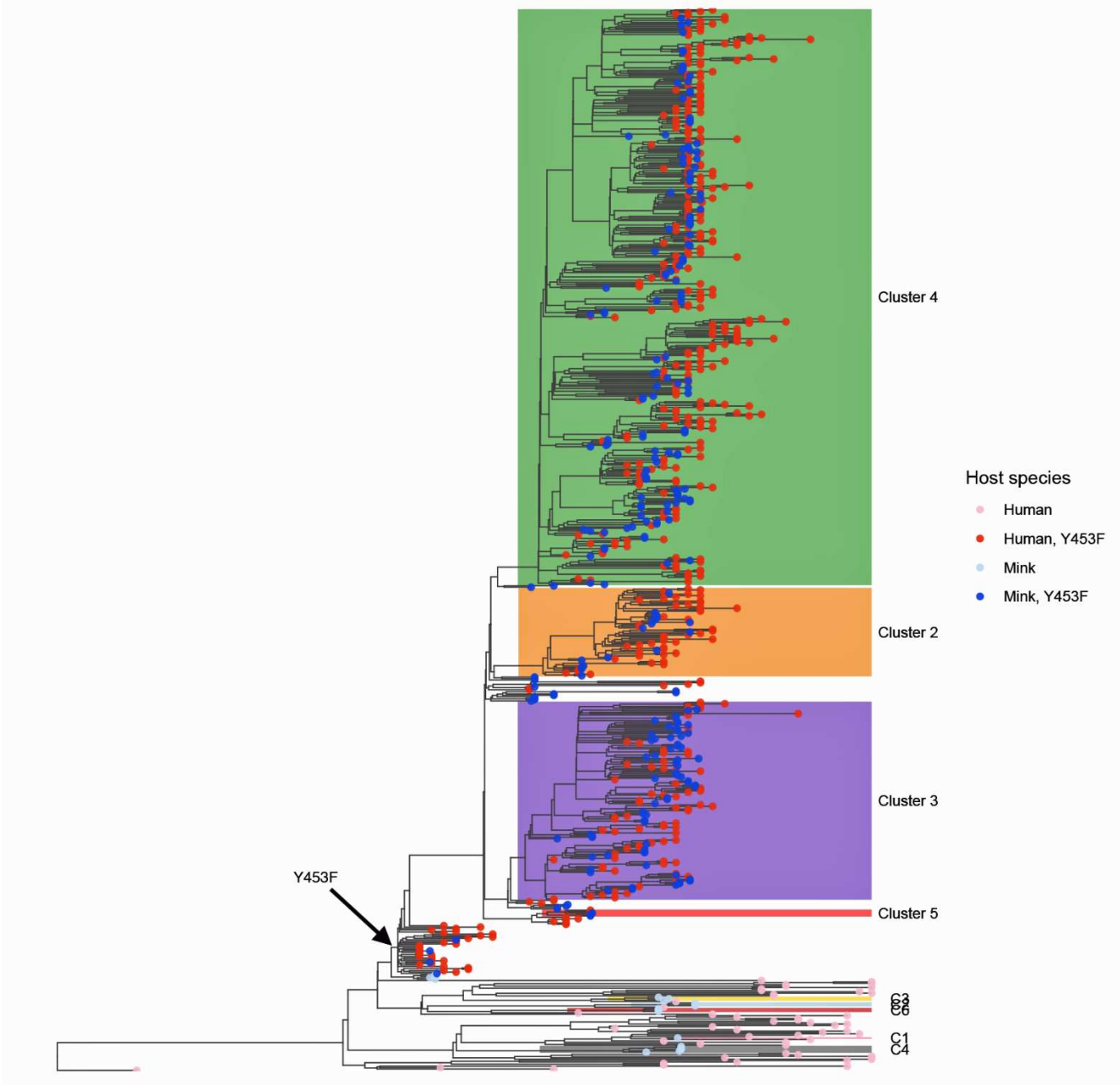


Figure 3. Phylogenetic tree based on whole-genome SARS-CoV-2 sequences from viruses obtained from humans and mink. Phylogenetic analysis was performed using BEAST2 with a strict clock model, GTR+gamma substitution model, and a BDSKY-serial tree prior. Shown here is the maximum clade-credibility (MCC) tree based on 17,500 post-burnin tree samples. Tips are colored based on host species (Human: red, Mink: blue), and on whether the encoded Spike protein contains the Y453F substitution (Yes: darker colors, No: lighter colors) resulting from the A22920T mutation. The Y453F substitution can be seen to evolve once (arrow pointing to tree branch), after which point it was retained in all descendant viruses. Also note how mink and human sequences are interspersed indicating frequent cross-species jumps.

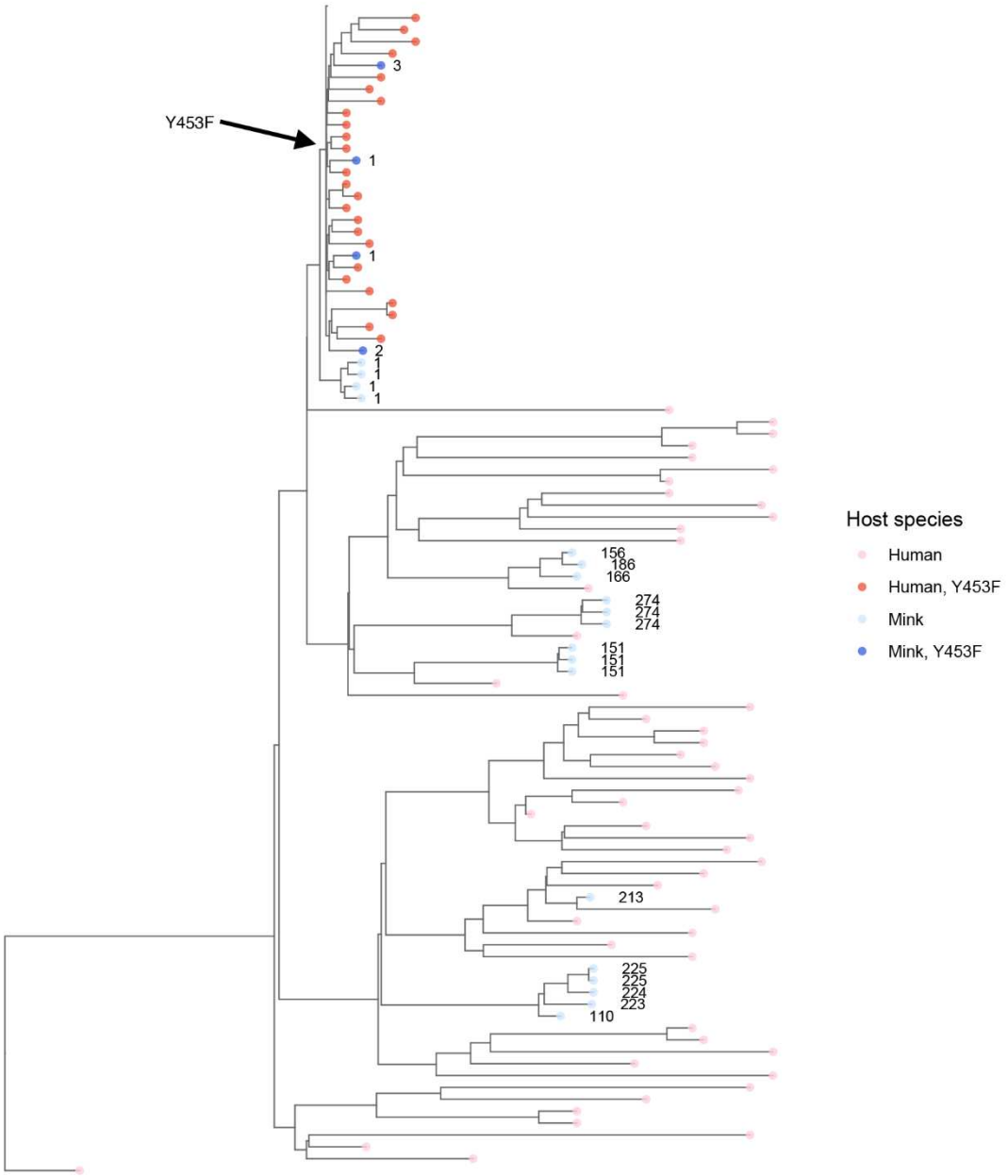


Figure 4. Zoom of phylogenetic tree from Figure 3 showing details around the branch where the Y453F S protein substitution occurred. Tips are colored based on host species (Human: red, Mink: blue) and on whether the encoded Spike protein contains the Y453F substitution (Yes: darker colors, no: lighter colors). Mink sequences are annotated with a number indicating the ID of the farm from which the sample was obtained. Note how only farm 1 had some mink without the Y453F change (light blue) and some with it (dark blue). This is consistent with the substitution occurring in the mink on farm 1.

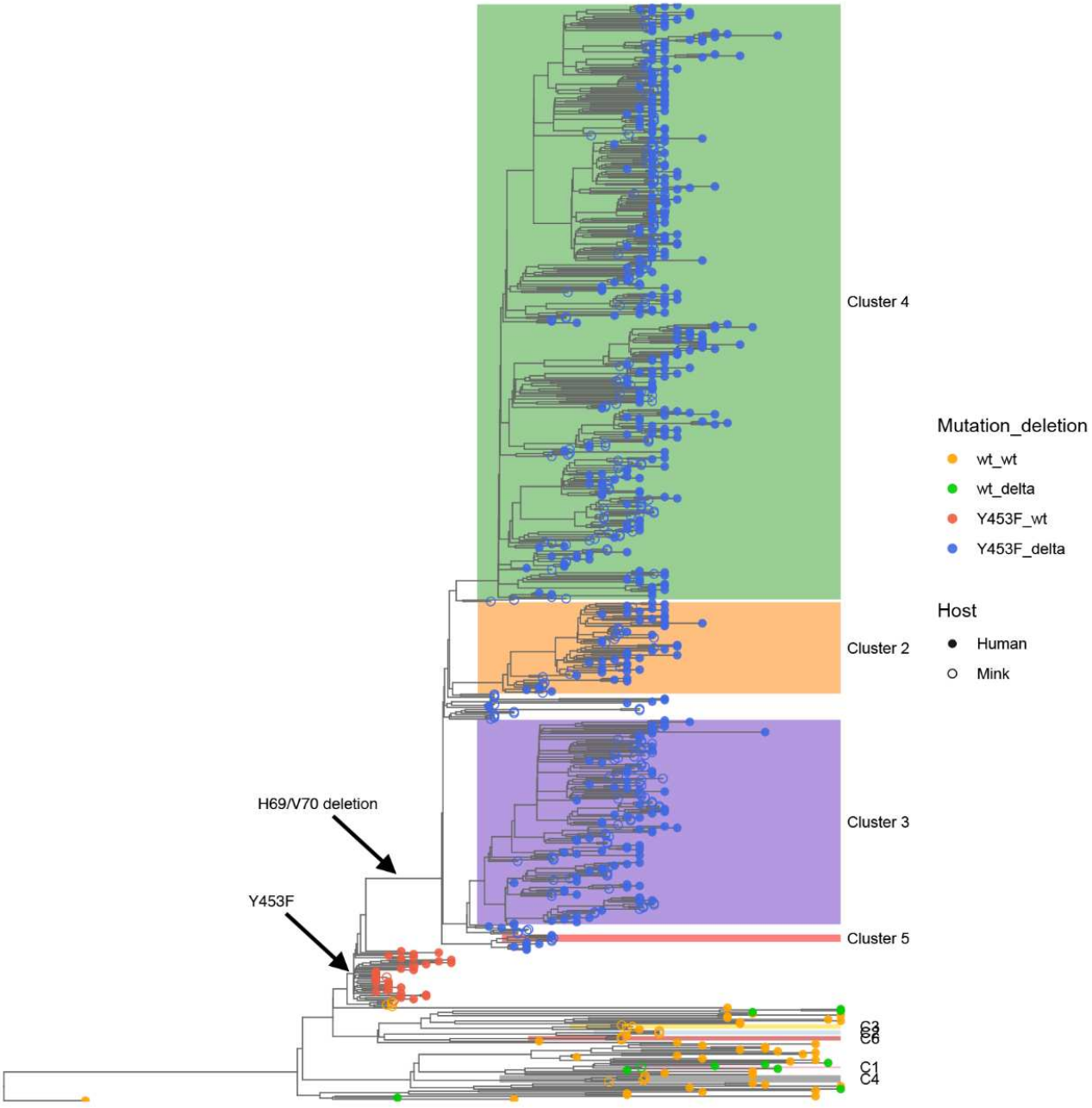


Figure 5. Phylogenetic tree from Figure 3 with tips colored according to presence or absence of the Y453F S protein substitution and the H69/V70 S protein deletion. The format used to label tips is <Y453F status>_<deletion status>, with "wt" indicating the absence of substitution or deletion, "Y453F" indicating the presence of that substitution, and "delta" indicating the presence of the deletion: wt_wt: orange, wt_delta: green, Y453F_wt: red, Y453F_delta: blue. Host species is indicated using open circles for mink and closed circles for human. Note that the H69/V70 deletion appears shortly after the Y453F substitution (arrows pointing to branches), and both changes are subsequently present in all descendant sampled viruses, from both humans and mink. The deletion was also present in 4 separate clades among viruses without Y453F (4 groups of green tips in bottom part of tree – see Figure 6 for further detail).

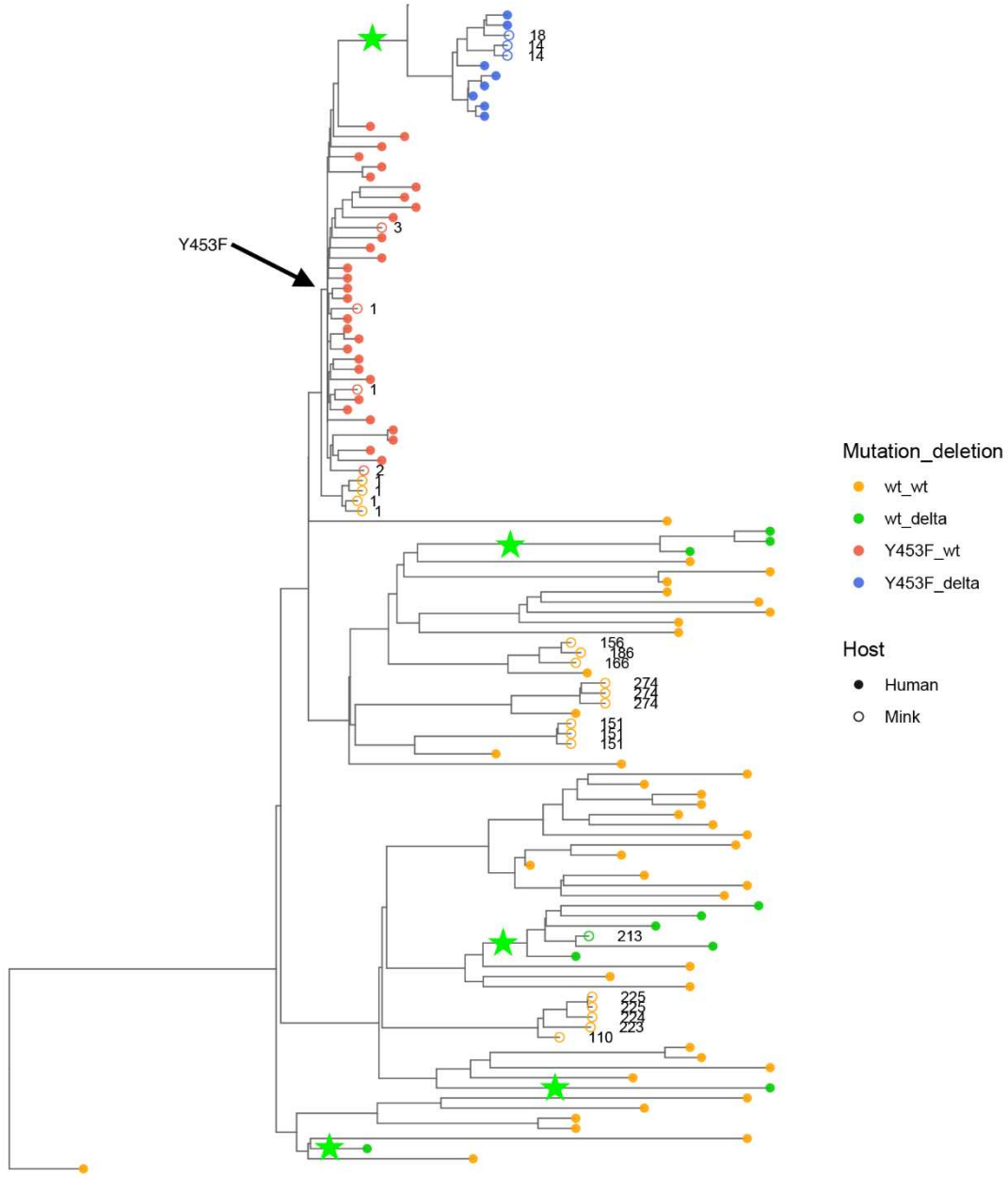


Figure 6. Zoom of phylogenetic tree from Figure 5 showing details around the branches where the H69/V70 deletion appeared. Color scheme is the same as in Figure 5. Mink-derived sequences are further annotated with a number indicating the farm ID from which the sample was obtained. The deletion can be seen to have evolved on a branch shortly after the Y453F substitution and to then have been retained in all viruses descending from this branch (see upper part of tree in Figure 5). Among the viruses, that do not have Y453F, the deletion is present in 4 separate clusters (green tips in bottom part of tree). The basal branches (where the deletions presumably evolved) of these 4 clusters are indicated with green stars. Two of the 4 clusters are human singletons (green closed circles near bottom of plot) and may correspond to independent introductions into DK of viruses already harboring the deletion. The two other clusters contain multiple sequences (3 and 6 respectively), indicating that the deletion may have evolved in DK and subsequently spread. One of these clusters contains only humans sequences, while the other contains the sequence from a single mink (from farm 213), that appears to have been infected by a human harboring virus with the deletion.

Cluster	Pangolin lineage	Clade	Spike protein signature [1]	Spike protein deletion*	No. of mink sequences	No. of farms	First sequence (date)	Last sequence (date)	Location
Initial cases	B.1.1.298	20B	(Y435F)	(H69/V70)	43	4	14-06	06-11	Northern Denmark
2	B.1.1.298	20B	Y453F	H69/V70	174	76	09-09	12-11	Northern/ Central Denmark
3	B.1.1.298	20B	Y453F	H69/V70	142	66	14-09	15-11	Northern/ Central Denmark
4	B.1.1.298	20B	Y453F	H69/V70	272	121	10-09	03-12	Northern/ Central Denmark
5	B.1.1.298	20B	Y453F	H69/V70	5	5	31-08	15-09	Northern Denmark
C1	B.1.258.9	20A	N439K; G1223S	H69/V70	2	1	03-11	-	Southern Denmark
C2	B.1.1.219	20B	F157L; (A845S)		13	4	16-10	13-11	Central Denmark
C3	B.1.1.170	20B	(G1167S)		6	3	23-10	29-10	Central Denmark
C4	B.1.536	20A	(N501T)		19	7	16-10	18-11	Southern Denmark
C6	B.1.1.294	20B			3	1	23-10	-	Northern Denmark
C7	B.1.1.159	20B			3	1	12-11		Southern Denmark
C8	B.1.177	20E	A222V		1	1	02-11	-	Northern Denmark

[1]: Changes shown in parenthesis were only found in some of the mink sequences in the cluster

Table 1. Summary of the SARS-CoV-2 sequences from infected mink from farms in DK. The features of the different Clusters (as identified in Figure 1) are shown. The later introductions into mink from infected humans, designated C1-C8 are also indicated; these viruses lack the Y453F change.