Manuscript version: Author's Accepted Manuscript
The version presented in WRAP is the author's accepted manuscript and may differ from the published version or Version of Record.

Persistent WRAP URL:
http://wrap.warwick.ac.uk/153884

How to cite:
Please refer to published version for the most recent bibliographic citation information.
If a published version is known of, the repository item page linked to above, will contain details on accessing it.

Copyright and reuse:
The Warwick Research Archive Portal (WRAP) makes this work by researchers of the University of Warwick available open access under the following conditions.
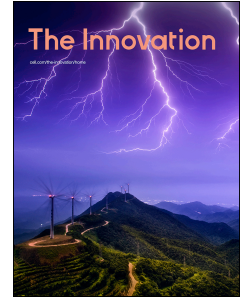
Publisher's statement:
Please refer to the repository item page, publisher's statement section, for further information.

For more information, please contact the WRAP Team at: wrap@warwick.ac.uk.

Journal Pre-proof

Assessing the extent of community spread caused by mink-derived SARS-CoV-2 variants

Liang Wang, Xavier Didelot, Yuhai Bi, George F. Gao

Please cite this article as: Wang, L., Didelot, X., Bi, Y., Gao, G.F, Assessing the extent of community spread caused by mink-derived SARS-CoV-2 variants, *The Innovation* (2021), doi: https://doi.org/10.1016/j.xinn.2021.100128.

PUBLIC SUMMARY

- SARS-CoV-2 transmission from human to mink is not lineage specific.

- Mink-derived SARS-CoV-2 variants keep human-to-human transmission.

- At least 12.5% of patients with mink-derived SARS-CoV-2 were caused by human-to-human transmission.

Assessing the extent of community spread caused by mink-derived SARS-CoV-2 variants

Liang Wang,[1, *] Xavier Didelot,[2] Yuhai Bi,[1,3,*] and George F Gao[1,3,*]

[1]CAS Key Laboratory of Pathogenic Microbiology and Immunology, Institute of Microbiology, Center for Influenza Research and Early-warning (CASCIRE), CAS-TWAS Center of Excellence for Emerging Infectious Diseases (CEEID), Chinese Academy of Sciences, Beijing 100101, China

[2]School of Life Sciences and Department of Statistics, University of Warwick, Coventry CV47AL, United Kingdom

[3]University of Chinese Academy of Sciences, Beijing 101408, China

* Correspondence: wangliang@im.ac.cn (L.W.); gaof@im.ac.cn (G.F.G); beeyh@im.ac.cn (Y.B)

ABSTRACT

SARS-CoV-2 has recently been found to have spread from humans to minks and then to have transmitted back to humans. However, it is unknown to what extent the human-to-human transmission caused by the variant has reached. Here, we used publicly available SARS-CoV-2 genomic sequences from both humans and minks collected in Denmark and Netherlands, and combined phylogenetic analysis with Bayesian inference under an epidemiological model to trace the possibility of person-to-person transmission. The results showed that at least 12.5% of all people being infected with dominated mink-derived SARS-CoV-2 variants in Denmark and Netherlands were caused by human-to-human transmission, indicating this "back-to-human" SARS-CoV-2 variant has already caused human-to-human transmission. Our study also indicated the need for monitoring this mink-derived and other animal source "back-to-human" SARS-CoV-2 in future and that prevention and control measures should be tailored to avoid large-scale community transmission caused by the virus jumped between animals and humans.

KEY WORDS：SARS-CoV-2；mink； human-to-human transmission

INTRODUCTION

Coronavirus disease 2019 (COVID-19) is caused by a novel type of coronavirus (known as SARS-CoV-2, 2019-nCoV, or HCoV-19)[1-3] that has led to more than 100 million infections of which at least 1.2 million have died worldwide on 10th November 2020, posing a global concern on public health.[4] Apart from humans, natural infection of SARS-CoV-2 have been found in several other species of

mammals through contact with COVID-19 patients, such as cats,[5] lions,[6] tigers,[6] dogs,[7] and minks.[8] Other animals have also been considered as possibly susceptible hosts (e.g. rabbit, pig, fox, mink, and civet) of SARS-CoV-2 through the entry test with pseudotype virus with S gene of SARS-CoV-2 and affinity abilities between the receptor binding domain (RBD) of S and host ACE2 protein.[9] In addition to human-to-animal transmission, SARS-CoV-2 in minks (*Neovison vison*) where it was initially introduced from humans could also transmit back to humans.[10] The virus was also shown to obtain some ongoing mink-adapted mutations such as Y453F, F486L, and N501T.[11] Since cross-species transmission has occurred and the SARS-CoV-2 can be transmitted back to humans from minks, it is important to clarify whether the "back-to-human" SARS-CoV-2 with ongoing mink-adapted mutations could further lead to transmission among humans. However, the reported study did not reach a conclusion on this point, but instead speculated that person-to-person transmission may have occurred.[10] Genomic sequence can be used to trace person-to-person transmission for SARS-CoV-2,[12] which represents an opportunity to confirm whether there was person-to-person transmission for the "back-to-human" SARS-CoV-2, even when epidemiological tracking information was not available or lacking. The main mink fur producing countries are Denmark, the Netherlands, Poland, and China.[13] It can be seen that Europe is the main production area of mink. Furthermore, mink fur delivered from European farms and sold at auction was worth 1.2 billion Euros in 2016.[14] Since tens of millions of minks have been culled to prevent further mutation and spread of the virus, the mink-derived SARS-CoV-2 variants (defined as those

isolated from minks) caused a catastrophic blow to the mink farming industry.

In this study, we used publicly available SARS-CoV-2 genomic sequences, and combined phylogenetic analysis with Bayesian inference under an epidemiological model to infer the probability of direct transmission between patients being infected with mink-derived SARS-CoV-2 variants in Denmark and Netherlands to evaluate the extent of human-to-human transmission caused by mink-derived SARS-CoV-2 variants.

MATERIALS AND METHODS

Data Collection, Filtration and Pre-processing

We retrieved genomic data from GISAID[15] on Jan 6 2021. We discarded genomic data with no exact collection date (accurate to days). Mink-derived sequences were defined as SARS-CoV-2 genomes isolated from minks. Since the most dominated mink-derived SARS-CoV-2 genomes belonged to B.1.298 and B.1.8 for Denmark and Netherlands, only human-derived and mink-derived SARS-CoV-2 genomes from these 2 lineages for Denmark and Netherlands were used. Genomic sequences were aligned using Mafft v7.310.[16] Then, we trimmed the uncertain regions in 3′ and 5′ terminals and also masked 30 sites (Supplementary Table 1) that are highly homoplastic and have no phylogenetic signal as previous noted (https://virological.org/t/issues-with-sars-cov-2-sequencing-data/473).

Phylogenetic Analysis

As recombination could impact the evolutionary signal, we searched for recombination events in these SARS-CoV-2 genomes using RDP4.[17] No evidence for recombination was found in our dataset. We used jModelTest v2.1.6[18] to find the best substitution model for each dataset from different countries according to the Bayesian Information Criterion. The best substitution model for datasets from Denmark and Netherlands was HKY and GTR+I, respectively. The list of genomic sequences used in this study were provided in Supplementary Table 2. The list of genomic sequences used in this study were openly shared via the GISAID initiative.[19] We then used the Bayesian Markov Chain Monte Carlo (MCMC) approach implemented in BEAST v1.10.4[20] to derive a dated phylogeny for SARS-CoV-2. Three replicate runs for each 100 million MCMC steps, sampling parameters and trees every 10,000 steps. As genomic sequences used in each dataset were all from the same lineage, we assumed that they followed a strict molecular clock. The estimation of the most appropriate coalescent models for Bayesian phylogenetic analysis was determined using both path-sampling and stepping-stone models.[21] The best fitting combination of prior of coalescent model was Bayesian skyline tree prior for both datasets (Supplementary Table 3). Tracer 1.7.1[22] was then used to check the convergence of MCMC chain (effective sample size >200) and to compute marginal posterior distributions of parameters, after discarding 10% of the MCMC chain as burn-in. We also reconstructed the host for each ancestral branch by using the Bayesian asymmetric discrete trait evolution model[23] under the Bayesian stochastic search variable selection

(BSSVS) framework. We determined whether there was sufficient temporal signal in each dataset, as it was the prerequisite for getting a reliable inference when performed phylodynamic analysis. Bayesian evaluation of temporal signal (BETS)[24] was used to evaluate the temporal signal in each dataset. BETS relies on the comparison of marginal likelihoods of two models: the heterochronous (with tip date) and isochronous (without tip date) models. Analyses were performed with at least three independent replicates of 100 million MCMC steps each, sampling parameters and trees every 10,000 steps with the best substitution model and most appropriate combination of molecular clock and coalescent models determined above for each dataset. The marginal likelihoods were estimated by PS. The Bayes factor (BF) was then calculated based on the likelihoods of two models (heterochronous and isochronous). If the log BF >5 (heterochronous model against isochronous model), it indicated there was sufficient temporal signal in this dataset. The log BF was estimated as 227 and 458 for datasets from Denmark and Netherlands, respectively, suggesting that the temporal signal was sufficiently strong. For convenience, we renumbered all mink-derived SARS-CoV-2 variants from humans (Supplementary Table 4).

Transmission Analysis

As viral genomes were incompletely sampled and the pandemic is currently ongoing, TransPhylo v1.4.4[25] was used to infer the transmission tree using the dated phylogeny generated above as input. The generation time (*i.e.* the time gap from

infection to onward transmission, denoted as G) of COVID-19 was previously estimated as 4.8±1.7 days,[26] and we used these values to compute the shape and scale parameter of a gamma distribution of G using the R package epitrix.[27] The distribution of sampling time (*i.e.* the time gap from infection to detection and sampling) was set equal to the distribution of generation time. We performed the TransPhylo analysis with at least 500,000 iterations simultaneously estimating the transmission tree, the proportion of sampling, the within-host coalescent time Neg, and the two parameters of the negative binomial offspring distribution (which represents the number of secondary cases caused by each infection). All results were generated after discarding the first part of the MCMC chains as burn-in (Supplementary Table 5). The MCMC mixing and convergence was assessed based on the effective sample size of each parameter (>200) and by visual examination of the MCMC traces. The probabilities of direct transmission from one host to another were estimated as the proportion of MCMC samples in which this direct transmission event occurred. The expected numbers of intermediates from one host to another were estimated as the average across the MCMC samples of the number of intermediates between the two hosts.

Evaluation of Phylogenetic Uncertainty on the Inference of Transmission Chain

Since the inference of transmission tree and further estimation on the probability of directed transmission were solely based on a dated-phylogeny. We then tested whether the uncertainty in phylogeny affected the result. Ten dated phylogenetic trees

were randomly selected from the MCMC chains for TransPhylo analysis. The parameter setting was the same as above. The estimated bidirectional probability of direct transmission for each patient pair was visualized by Cytoscape v3.8.2.[28]

RESULTS

The Geographical and Phylogenetic Distribution of Mink-derived SARS-CoV-2 Variants

As of Jan 6 2021, there were a total of 761 mink-derived SARS-CoV-2 genomes have been available. They came from four countries: Canada (4 genomes), Denmark (454 genomes), Netherlands (291 genomes), and Poland (12 genomes), respectively. All these mink-derived SARS-CoV-2 genomes belonged to 15 lineages. For viruses from Canada and Poland, they all belonged to lineage B.1.1 (Figure 1). In Netherlands, the dominated strains came from lineage B.1.298 (Figure 1). However, the dominated mink-derived variants in Denmark belonged to lineage B.1.8 (Figure 1). The widely phylogenetic distribution of mink-derived SARS-CoV-2 genomes suggested cross-species transmission events of SARS-CoV-2 from human to mink were not lineage specific. Considering the number of mink-derived variants, further analysis was mainly focused on those from lineage B.1.8 in Denmark and lineage B.1.298 in Netherlands.

Identification of Cross-species Transmission Events

We used a discrete trait analysis to infer the ancestral host for each branch. An

independent cross-species transmission event was considered to be occurred only if a clade meets all the following criteria: i) the direct two branches after the root of the clade have different host, ii) posterior probability of both branch and ancestral host for the root of the clade > 0.8. In Denmark dataset, we found three independent cross-species transmission events (Figure 2). All of them were caused by human-to-mink transmission. In addition, we also found six SARS-CoV-2 genomes (in Clade I) from human were closed to mink-derived viral genomes, indicating they were highly likely to be transmitted from mink to human. However, we could not determine how many independent cross-species transmission events occurred due to the low posterior probability of branches. In Netherlands, three independent cross-species transmission events occurred in lineage B.1.298 (Figure 3). One of them was transmitted from human to mink, other two events were caused by transmission from mink to human, which contained one and five cases, respectively. Besides, we also found a cluster denoted as Clade I contained nine SARS-CoV-2 genomes from human were closed to mink-derived viral genomes, indicating they were highly likely to be transmitted from mink to human. However, we could only be sure that at least one independent cross-species transmission event has occurred between them. Besides, we also found additional four SARS-CoV-2 genomes from human were scattered within Clade III, indicating these four patients were also infected with mink-derived SARS-CoV-2 variants. Totally, we identified 18 patients being infected with mink-derived SARS-CoV-2 variants. We further tested whether there were human-to-human transmission events in those who infected with mink-derived

SARS-COV-2 variants.

Inference of Person-to-person Transmission Events of Mink-derived SARS-CoV-2 Variants

We then calculated the probability of direct transmission between humans infected with mink-derived SARS-CoV-2 variants. In order to reduce the calculation, only clades with highly posterior probability of its root and contained humans infected with mink-derived SARS-CoV-2 variants were used to further analysis. In Denmark dataset, there are three patient pairs (D2/D3, D5/D6, and D1/D3) with bidirectional probability of direct transmission >0.5 (0.998, 0.731, and 0.607, respectively) (Figure 4A). Meanwhile, the number of intermediates between D2/D3, D5/D6, and D1/D3 were estimated as 0.002, 0.271, and 0.412, respectively (Figure 4B). All these results suggested that these three patient pairs were more likely to be transmitted from each other directly. In the Dutch dataset, we also found 2 pair of patients in Clade I (N7/N8, and N3/N4) with bidirectional probability of direct transmission >0.5 (0.95, and 0.931, respectively) (Figure 4C). In addition, the number of intermediates between N7/N8, and N3/N4 were estimated as 0.05, and 0.069, respectively (Figure 4D). There are also 2 pairs of patients in Clade II (N10/N11 and N13/N14) with bidirectional probability of direct transmission >0.5 (0.989, and 0.978, respectively) (Figure 4E). In addition, the number of intermediates between N10/N11, and N13/N14 were estimated as 0.011, and 0.022, respectively (Figure 4F).

Validation of the Direct Transmission Events

Since limited variations detected in mink-derived SARS-CoV-2 variants could not result in a highly resolved phylogeny, we next wanted to test how did phylogenetic uncertainty affect the result by repeating the analysis based on 10 trees randomly selected from MCMC chain. In the Danish dataset, the cluster with D1, D2 and D3 always contained a patient pair with highly bidirectional probability of direct transmission (Figure 5). However, the bidirectional probability of direct transmission for D5 and D6 was lower than 0.5 in 2 randomly selected trees, indicating the inference of direct transmission between D5 and D6 could be affected by the phylogenetic uncertainty (Figure 5). In this case, we could conclude that only one person-to-person transmission event occurred in the Danish dataset. For the Dutch dataset, we found a similar pattern as in the Danish dataset that the phylogenetic uncertainty highly affected the inference of who infect whom. However, there was at least one direct transmission event with high bidirectional probability occurred in each cluster for 10 randomly selected phylogenies (Figure 6). Besides, we also found that N10-N14 are all employees in the same mink farm, indicating that the direct transmission between them could be more likely occurred. In this case, we could conclude that at least two person-to-person transmission events occurred in the Dutch dataset. In summary, we totally identified at least 3 direct transmission events with high bidirectional probability among humans infected with mink-derived SARS-CoV-2 variants in Denmark and Netherlands. It accounted for 12.5% of all people infected with mink-derived SARS-CoV-2 variants in this study.

We also found some mutations arising in those "back-to-human" SARS-CoV-2 genomes compared to their close-related mink-derived ones. In the Danish dataset, C2062T (locating at the 5' terminal of SARS-CoV-2 genome) was detected in D4. However, this mutation was not detected in other close related human SARS-CoV-2 and closest mink-derived variant. A nonsynonymous mutation (C12008T result in Leu3915Phe in ORF1ab) were lost in both D5 and D6, compared to their closest related mink-derived variant. In the Dutch dataset, more mutations were detected. Among them, we found that there was no common mutation shared by all human genomes. Together with the limited number of mink-derived SARS-CoV-2 genomes from human, we are currently unable to determine whether there are human adaptive mutations.

DISCUSSION

Since the SARS-CoV-2 carried by mink could be transmitted back to humans,[10] it had led to the mass culling of infected animals, posing a huge threat to the public health and economy. The first thing we need to evaluate is whether the variant can continue to spread from person to person, and to study the extent of the current human-to-human transmission. We found that the phylogenetic type of dominant strains in different countries were not consistent (Figure 1), indicating that the cross-species transmission events of SARS-CoV-2 from human to mink were not lineage specific. In other words, many phylogenetic subtypes of SARS-CoV-2 can be

transmitted from human to mink. However, if all subtypes of the SARS-CoV-2 can be transmitted to mink is still unknown and needs further research and confirmation. Several independent cross-species transmission events were identified in this study, which contained both human-to-mink and mink-to-human direction. We also detected at least three human-to-human transmission events with highly bidirectional probability in this study. However, we are not sure who infected whom, mainly due to the phylogenetic uncertainty caused by limited mutations. The phylogenetic uncertainty also caused different number of direct transmission event for each dataset (Figure 5&6). Yet, there was always one direct transmission event with high bidirectional probability occurred in each dataset. Under these circumstances, we could conclude that there were at least three direct transmission events identified in Denmark and Netherlands, accounting for at least 12.5% of all people infected with mink-derived SARS-CoV-2 variants in this study. However, the extent of human-to-human transmission caused by mink-derived SARS-CoV-2 variants was considered to be underestimated. The reasons are summarized as follows. First, not all viral genomes of patients infected with mink-derived SARS-CoV-2 variants were available right now. Second, the criteria of identifying direct transmission event were strict in this study, leading to a low true positive ratio. In this case, the mink-derived SARS-CoV-2 variants in human and then the extent of person-to-person transmission caused by this variant should be continuously monitoring. Despite mink was the only species so far that could be easily infected by humans with SAR-CoV-2, and then spill the mutants back to humans again, this phenomenon might also exist in other

non-human mammals which could be infected by SARS-CoV-2 and were in frequent contact with humans. Under these circumstances, the contact between humans and susceptible animals should be cautious to prevent humans from transmitting SARS-CoV-2 to animals, so as to prevent the virus from continuously circulating and evolving in the animals. This will not only minimize the impact of the SARS-CoV-2 on the breeding industry, as increased mortality was detected in farmed minks that were positive to SARS-CoV-2 RNA,[29] but also decrease the probability of generating some novel and unpredictable mutants of SARS-CoV-2 within animals, thereby threatening public health.

REFERENCES

1.   Zhu, N., Zhang, D., Wang, W., et al. (2020). A novel coronavirus from patients with pneumonia in China, 2019. N Engl J Med. *382*, 727-733.

2.   Jiang, S.B., Shi, Z.L., Shu, Y.L., et al. (2020). A distinct name is needed for the new coronavirus. Lancet. *395*, 949.

3.   Tan, W., Zhao, X., Ma, X., et al. (2020). A novel coronavirus genome identified in a cluster of pneumonia cases — Wuhan, China 2019−2020. China CDC Weekly. *2*, 61-62.

4.   Wang, C., Horby, P.W., Hayden, F.G., et al. (2020). A novel coronavirus outbreak of global health concern. Lancet. *395*, 496.

5.   Garigliany, M., Van, L. A.S., Clercx, C., et al. (2020). SARS-CoV-2 natural transmission from human to cat, Belgium, March 2020. Emerg Infect Dis. *26*, 3069-3071.

6.   McAloose, D., Laverack, M., Wang, L.Y., et al. (2020). From people to panthera: natural SARS-CoV-2 infection in tigers and lions at the Bronx Zoo. Mbio. *11*.

7.   Sit, T.H.C., Brackman, C.J., Ip, S.M., et al. (2020). Infection of dogs with SARS-CoV-2. Nature. *586*, 776-778.

8.   Hammer, A.S., Quaade, M.L., Rasmussen, T.B., et al. (2020). SARS-CoV-2 transmission between mink (Neovison vison) and humans, Denmark. Emerg Infect Dis. *27*, 547-551.

9.   Wu, L., Chen, Q., Liu, K., et al. (2020). Broad host range of SARS-CoV-2 and the molecular basis for SARS-CoV-2 binding to cat ACE2. Cell Discov. *6*, 68.

10.   Oude, M.B.B., Sikkema, R.S., Nieuwenhuijse D.F., et al. (2020). Transmission of SARS-CoV-2 on mink farms between humans and mink and back to humans. Science. *8*, 172-177.

11. van, D.L., Tan, C.C.S., Lam, S.D., et al. (2020). Recurrent mutations in SARS-CoV-2 genomes isolated from mink point to rapid host-adaptation. bioRxiv. 2020.11.16.384743.

12. Wang, L., Didelot, X., Yang, J., et al. (2020). Inference of person-to-person transmission of COVID-19 reveals hidden super-spreading events during the early outbreak phase. Nat Commun. *11*, 5006.

13. Fenollar, F., Mediannikov, O., Maurin, M., et al. (2021). Mink, SARS-CoV-2, and the human-animal interface. Front Microbiol. *12*, 663815.

14. Hansen, H.O. (2017). European mink industry - socio-economic impact assessment. https://www.altinget.dk/misc/Fur-Invasive-19-09.pdf.

15. Shu, Y.L., McCauley, J. (2017). GISAID: global initiative on sharing all influenza data - from vision to reality. Eurosurveillance. *22*, 2-4.

16. Katoh, K., Misawa, K., Kuma, K., et al. (2002). MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res. *30*, 3059-3066.

17. Martin, D.P., Murrell, B., Golden, M., et al. (2015). RDP4: detection and analysis of recombination patterns in virus genomes. Virus Evol. *1*, vev003.

18. Darriba, D., Taboada, G.L., Doallo, R., et al. (2012). jModelTest 2: more models, new heuristics and parallel computing. Nat Methods. *9*, 772.

19. Elbe, S., Buckland-Merrett, G. (2017). Data, disease and diplomacy: GISAID's innovative contribution to global health. Global Challenges. *1*, 33-46.

20. Suchard, M.A., Lemey, P., Baele, G., et al. (2018). Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. Virus Evol. *4*, vey016.

21. Baele, G., Li, W.L., Drummond, A.J., et al. (2013). Accurate model selection of relaxed molecular clocks in bayesian phylogenetics. Mol Biol Evol. *30*, 239-243.

22. Rambaut, A., Drummond, A.J., Xie, D., et al. (2018). Posterior summarization in Bayesian phylogenetics using Tracer 1.7. Syst Biol. *67*, 901-904.

23. Lemey, P., Rambaut, A., Drummond, A.J., et al. (2009). Bayesian phylogeography finds its roots. PLoS Comput Biol. *5*, e1000520.

24. Duchene, S., Lemey, P., Stadler, T., et al. (2020). Bayesian evaluation of temporal signal in measurably evolving populations. Mol Biol Evol. *37*, 3363-3379.

25. Didelot, X., Fraser, C., Gardy, J., et al. (2017). Genomic infectious disease epidemiology in partially sampled and ongoing outbreaks. Mol Biol Evol. *34*, 997-1007.

26. Challen, R., Brooks-Pollock, E., Tsaneva-Atanasova, K., et al. (2020). Meta-analysis of the SARS-CoV-2 serial interval and the impact of parameter uncertainty on the COVID-19 reproduction number. medRxiv. 2020.11.17.20231548.

27. Thibaut, J., Anne, C. (2019). epitrix: small helpers and tricks for epidemics analysis. https://CRAN.R-project.org/package=epitrix.

28. Shannon, P., Markiel, A., Ozier, O., et al. (2003). Cytoscape: A software environment for integrated models of biomolecular interaction networks. Genome Res. *13*, 2498-2504.

29. Oreshkova, N., Molenaar, R.J., Vreman, S., et al. (2020). SARS-CoV-2 infection in farmed minks, the Netherlands, April and May 2020. Euro Surveill. *25*, 2-8.

ACKNOWLEDGMENTS

AUTHOR CONTRIBUTIONS

L.W., Y.B., and G.F.G designed and coordinated the study. L.W. collected data and performed the analysis. L.W., X.D., Y.B., G.F.G contributed to the critical interpretation of the results. L.W. wrote the paper. L.W., X.D., Y.B., and G.F.G., revised the manuscript. All authors reviewed and edited the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

SUPPLEMENTAL INFORMATION

Supplemental Information contains five tables.

LEAD CONTACT WEBSITE

http://www.caspmi.cn/gaog/index.html

Figure Legends

Figure 1. The Geographical and Phylogenetic Distribution of Mink-derived SARS-CoV-2 Variants. The size of the pie is proportional to the number of mink-derived SARS-CoV-2 genomes available in GISAID. The different colors indicate the different subtype of mink-derived SARS-CoV-2 variants.

Figure 2. Identification of Cross-species Transmission Events in Danish Dataset.Maximum clade credibility phylogeny for the Danish dataset. Branch colors indicate the most probable ancestral host. For selected nodes, colored numbers show the posterior probabilities of ancestral host and numbers in black are clade posterior probabilities.

Figure 3. Identification of Cross-species Transmission Events in the Dutch Dataset. Maximum clade credibility phylogeny for the Dutch dataset. Branch colors indicate the most probable ancestral host. For selected nodes, colored numbers show the posterior probabilities of ancestral host and numbers in black are clade posterior probabilities.

Figure 4. Identification of Direct Transmission Between Humans Being Infected with Mink-derived SARS-CoV-2 Variants.

(A) The probability of directed transmission from infector (row) to infectee (column) for the Danish dataset.

(B) The number of intermediates in the transmission chain between each pair for the Danish dataset.

(C) The probability of directed transmission from infector (row) to infectee (column)

for the Clade I in the Dutch dataset.

(D) The number of intermediates in the transmission chain between each pair for Clade I in the Dutch dataset.

(E) The probability of directed transmission from infector (row) to infectee (column) for the Clade II in the Dutch dataset.

(F) The number of intermediates in the transmission chain between each pair for Clade II in the Dutch dataset.

Figure 5. The Influence of Phylogenetic Uncertainty on the Inference of Direct Transmission Pair for Danish Dataset.The direct transmission pair for the Danish dataset estimated from 10 randomly selected trees from MCMC chains. Nodes represent patients and the lines between each node represents the direct transmission. The width of the line is proportional to the bidirectional probability of direct transmission between each patient. If the bidirectional probability of direct transmission between each patient was >0.5, the line is purple, otherwise it is black.

Figure 6. The Influence of Phylogenetic Uncertainty on the Inference of Direct Transmission Pair for the Dutch Dataset. The network graph of direction transmission pair for the Dutch dataset estimated from 10 randomly selected trees from MCMC chains. Nodes represent patients and the lines between each node represents the direct transmission. The yellow and green nodes represent nodes from Cluster I and II, respectively. The width of the line is proportional to the bidirectional probability of direct transmission between each patient. If the bidirectional probability of direct transmission between each patient was >0.5, the line is purple, otherwise it is black.