Highly divergent white-tailed deer SARS-CoV-2 with potential deer-to-human transmission

Bradley Pickering[1,9,10 *], Oliver Lung[1,11 *], Finlay Maguire[2,12,13 *], Peter Kruczkiewicz[1], Jonathan D. Kotwa[3], Tore Buchanan[4], Marianne Gagnier[5], Jennifer L. Guthrie[6,16], Claire M. Jardine[7], Alex Marchand-Austin[6], Ariane Massé[5], Heather McClinchey[8], Kuganya Nirmalarajah[3], Patryk Aftanas[13], Juliette Blais-Savoie[3], Hsien-Yao Chee[3], Emily Chien[3], Winfield Yim[3], Melissa Goolia[1], Matthew Suderman[1], Mathieu Pinette[1], Greg Smith[1], Daniel Sullivan[1,11], Jossip Rudar[1], Elizabeth Adey[4], Michelle Nebroski[1], Marceline Côté[17,18,19], Geneviève Laroche[17,18,19], Allison J. McGeer[14, 20], Larissa Nituch[4], Samira Mubareka[3,14+], Jeff Bowman[4,15+]

[1]National Centre for Foreign Animal Disease, Canadian Food Inspection Agency, Winnipeg, Manitoba, Canada

[2]Faculty of Computer Science, Dalhousie University, Nova Scotia, Canada

[3]Sunnybrook Research Institute, Toronto, Ontario, Canada

[4]Wildlife Research and Monitoring Section, Ontario Ministry of Northern Development, Mines, Natural Resources and Forestry, Peterborough, Ontario, Canada

[5]Ministère des Forêts, de la Faune et des Parcs, Québec City, Québec, Canada

[6]Public Health Ontario, Toronto, Ontario, Canada

[7]Canadian Wildlife Health Cooperative, Ontario-Nunavut, Department of Pathobiology, University of Guelph, Guelph, Ontario, Canada

[8]Public Health, Health Protection and Surveillance Policy and Programs Branch, Ontario Ministry of Health

[9]Department of Veterinary Microbiology and Preventative Medicine, College of Veterinary Medicine, Iowa State University, Ames, Iowa, USA

[10]Department of Medical Microbiology and Infectious Diseases, University of Manitoba, Winnipeg

[11]Department of Biological Sciences, University of Manitoba, Winnipeg, Manitoba, Canada

[12]Department of Community Health & Epidemiology, Faculty of Medicine, Dalhousie University, Nova Scotia, Canada

[13]Shared Hospital Laboratory, Toronto, Ontario, Canada

[14]Department of Laboratory Medicine and Pathobiology, University of Toronto, Toronto, Ontario, Canada

[15]Environmental and Life Sciences Graduate Program, Trent University, Ontario, Canada

[16] Department of Microbiology & Immunology, Western University, London, Ontario, Canada

[17]Department of Biochemistry, Microbiology and Immunology, University of Ottawa, Ottawa, Ontario, Canada

[18]Ottawa Institute of Systems Biology, University of Ottawa, Ottawa, Ontario, Canada

[19]Centre for Infection, Immunity, and Inflammation, University of Ottawa, Ottawa, Ontario, Canada

[20]Sinai Health System, Toronto, Ontario, Canada

* Authors contributed equally to the work

+ Authors jointly supervised the work

Co-corresponding authors: bradley.pickering@canada.ca (B. Pickering); samira.mubareka@sunnybrook.ca (S. Mubareka); jeff.bowman@ontario.ca (J. Bowman).

Summary

Wildlife reservoirs of SARS-CoV-2 can lead to viral adaptation and spillback from wildlife to humans (Oude Munnink et al., 2021). In North America, there is evidence of spillover of SARS-CoV-2 from humans to white-tailed deer (*Odocoileus virginianus*), but no evidence of transmission from deer to humans (Hale et al., 2021; Kotwa et al., 2022; Kuchipudi et al., 2021). Through a multidisciplinary research collaboration for SARS-CoV-2 surveillance in Canadian wildlife, we identified a new and highly divergent lineage of SARS-CoV-2. This lineage has 76 consensus mutations including 37 previously associated with non-human animal hosts, 23 of which were not previously reported in deer. There were also mutational signatures of host adaptation under neutral selection. Phylogenetic analysis revealed an epidemiologically linked human case from the same geographic region and sampling period. Together, our findings represent the first evidence of a highly divergent lineage of SARS-CoV-2 in white-tailed deer and of deer-to-human transmission.

Main

High consequence coronaviruses, including severe acute respiratory syndrome coronavirus (SARS-CoV), SARS-CoV-2, and Middle East respiratory syndrome coronavirus (MERS-CoV) have putative animal origins and are likely transmitted to humans either directly from reservoir hosts or through intermediates such as civets or camels (Andersen et al., 2020; Boni et al., 2020; Cui et al., 2019; Haagmans et al., 2014; Hu et al., 2021; Memish et al., 2013). This may be followed by sustained human-to-human transmission with ongoing viral adaptation, as we have seen with SARS-CoV-2 and the emergence of variants of concern (VOCs). Additional viral diversity may be gained through inter-species viral transmission, as was observed during human-mink-human transmission of SARS-CoV-2 (Lu et al., 2021). The emergence of divergent viral lineages can have considerable impact on viral immunology, biology and epidemiology (e.g., host immune evasion, disease severity, and transmission), impacting individual and population health. The Omicron VOC is a highly divergent lineage having amassed 59 mutations including 37 in its spike protein, which has had global impacts on healthcare systems and societies (Mannar et al., 2022). There are several competing theories behind the origin of this VOC; however, transmission analysis of acquired mutations and molecular docking experiments indicate it may have originated after transmission from humans to mice and back (i.e., spillback) (Wei et al., 2021) or another, as of yet unidentified, animal reservoir. Establishment of an animal reservoir of SARS-CoV-2 through viral persistence within a susceptible species may lead to repeated spillback events into the human population, with the risk of sustained human-to-human transmission (Hallmaier-Wacker et al., 2017).

As of January 2022, SARS-CoV-2 has been shown to infect at least 29 non-human mammalian species through both observational and experimental studies in free-living, captive, domestic, and farmed animals (Abdel-Moneim & Abdelwhab, 2020; Shu & McCauley, 2017; Tan et al., 2022). Homology of the primary SARS-CoV-2 host cell receptor, human angiotensin converting enzyme 2 (hACE2) among a broad range of animal species as predicted through *in silico* modelling (Damas et al., 2020) may be a key host determinant. Zooanthroponosis has been documented in outbreaks of SARS-CoV-2 among farmed (and escaped or feral) mink (*Neovison vison*) in Europe and North America (Molenaar et al., 2020; Shriner et al., 2021) and pet hamsters (*Mesocricetus auratus*) (Yen et al., 2022). Notably, the Netherlands experienced outbreaks of SARS-CoV-2 in mink farms (Oude Munnink et al., 2021), and whole genome sequencing (WGS) provided evidence for the emergence of a "cluster 5" variant among farmed mink with a unique combination of mutations, and spillover from mink to humans (Oude Munnink et al., 2021). These mutations raised concerns around vaccine efficacy contributing to the decision in Denmark to depopulate mink (Frutos & Devaux, 2020). The finding of SARS-CoV-2 in pet hamsters led Hong Kong authorities to cull thousands of animals (Pang & Siu, 2022).

There have been suggestions (based on experimental data for VOCs) that SARS-CoV-2 host cell receptor tropism has expanded over time, increasing concerns about the potential for spillover into animals. For example, the Alpha variant is capable of infecting mice (*Mus musculus*), and the Omicron variant spike protein has been shown to bind avian ACE2 receptors, whereas ancestral SARS-CoV-2 did not (Peacock et al., 2022; Shuai et al., 2021).

The white-tailed deer (*Odocoileus virginianus*) is a common and widespread ungulate in North America and has recently been demonstrated to be susceptible to SARS-CoV-2. An experimental study first showed that deer developed subclinical infection (Palmer et al., 2021). A subsequent study found that 40% of free-ranging deer sampled in Michigan, Illinois, New York, and Pennsylvania, USA were positive for antibodies to SARS-CoV-2 (Chandler et al., 2021). Sylvatic transmission among deer and multiple spillovers from humans to deer have also been confirmed (Hale et al., 2021; Kotwa et al., 2022; Kuchipudi et al., 2021). To date, SARS-CoV-2 viruses observed in deer have appeared similar to prevalent lineages circulating in nearby human populations, suggesting multiple, recent spillover events (Kotwa et al., 2022; Kuchipudi et al., 2021). There is no evidence, so far, of deer-to-human transmission of SARS-CoV-2.

In response to evidence of free-ranging white-tailed deer (WTD) infection with SARS-CoV-2, the potential establishment of a deer reservoir, and the risk of deer-to-human transmission, we initiated a SARS-CoV-2 surveillance program of WTD in Ontario, Canada to better understand SARS-CoV-2 prevalence in regional WTD. This work provides evidence of parallel evolution of SARS-CoV-2 in a deer population in Southwestern Ontario with unsustained deer-to-human transmission.

Highly divergent SARS-CoV-2 found in deer

From 1 November to 31 December 2021, 300 WTD were sampled from Southwestern (N=249, 83%) and Eastern (N=51, 17%) Ontario, Canada during the annual hunting season (Figure 1). The majority of sampled WTD were adults (94%) with comparable numbers of females (N=135, 45%) and males (N=165, 55%). We collected 213 nasal swabs and tissue from 294 retropharyngeal lymph nodes (RPLN) which were tested for the presence of SARS-CoV-2 RNA using RT-PCR.

Five of 213 (2.3%) nasal swabs were positive by two independent RT-PCR analyses at separate institutes (UTR and E gene Ct<40; and E and N2 gene Ct<36). Sixteen RPLN were confirmed by PCR. Overall, SARS-CoV-2 RNA was detected in 21 samples representing 6% (17/298) of hunter-harvested WTD included in the

present study; all positive animals were adult deer from Southwestern Ontario and the majority (65%) were female (Figure 1, Table S1). Two deer were excluded due to indeterminate RPLN results with no corresponding nasal swab.

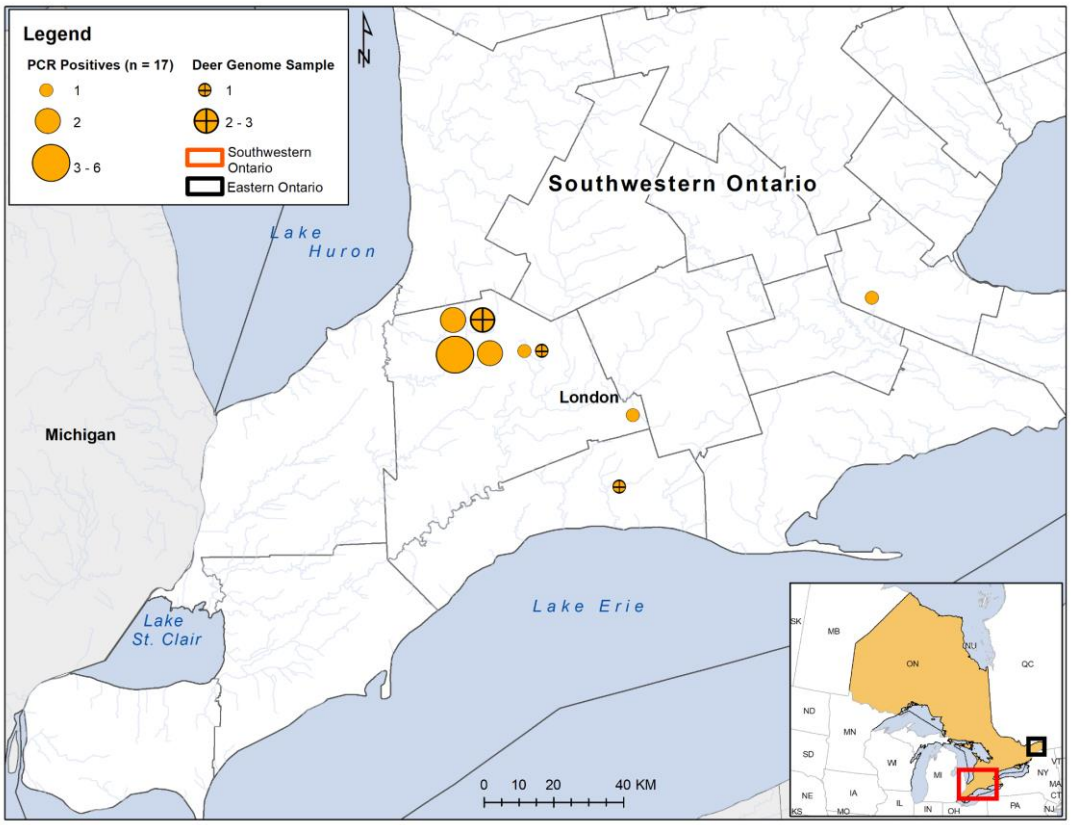

Figure 1: **SARS-CoV-2 RNA detection in WTD sampled in Southwestern and Eastern Ontario in 2021. Circle size indicates the relative number of positive WTD (n=17/298), with crosses showing samples from which high quality viral genomes were recovered (n=5). The detailed map depicts Southwestern Ontario (the red rectangle on the inset map). SARS-CoV-2 RNA was not detected in samples from Eastern Ontario.**

From the 5 nasal swabs, 3 high-quality SARS-CoV-2 consensus genomes were recovered using a standard amplicon-based approach. All samples were also independently extracted and sequenced using a capture-probe-based approach for confirmation. By combining the amplicon and capture-probe sequencing data, 5 high quality genomes (WTD nasal swabs: 4581, 4645, 4649, 4658, and 4662) and 2 partial genomes (WTD RPLNs: 4538, 4534) were recovered (deposited accessions can be found in Table S1). Human RNAse P PCR was negative for these samples and the majority (65-79%) of non-SARS-CoV-2 amplicon reads mapped to the WTD reference genome, demonstrating that contamination from human-derived SARS-CoV-2 sequences was highly unlikely.

Maximum-likelihood and parsimony-based phylogenetic analysis showed these WTD genomes formed a highly divergent clade within the B.1 PANGO lineage/20C Nextstrain clade (100% Ultrafast Bootstrap [UFB]). This B.1 lineage encompasses significant diversity and was the underlying backbone from which the Beta VOC, Epsilon and Iota Variants Under Investigation (VUIs), and significant mink (*Neovison*) outbreaks emerged (Figure 2). The WTD clade forms a very long branch with 76 conserved nucleotide mutations relative to ancestral SARS-CoV-2 (Wuhan Hu-1) and 49 relative to their closest common ancestor with other

genomes in GISAID (as of February 2022). The closest branching genomes in GISAID are human-derived sequences from Michigan, USA, sampled approximately 1 year prior (November/December 2020). These sequences in turn are closely related to a mixed clade of human and mink samples from Michigan collected in September/October 2020.

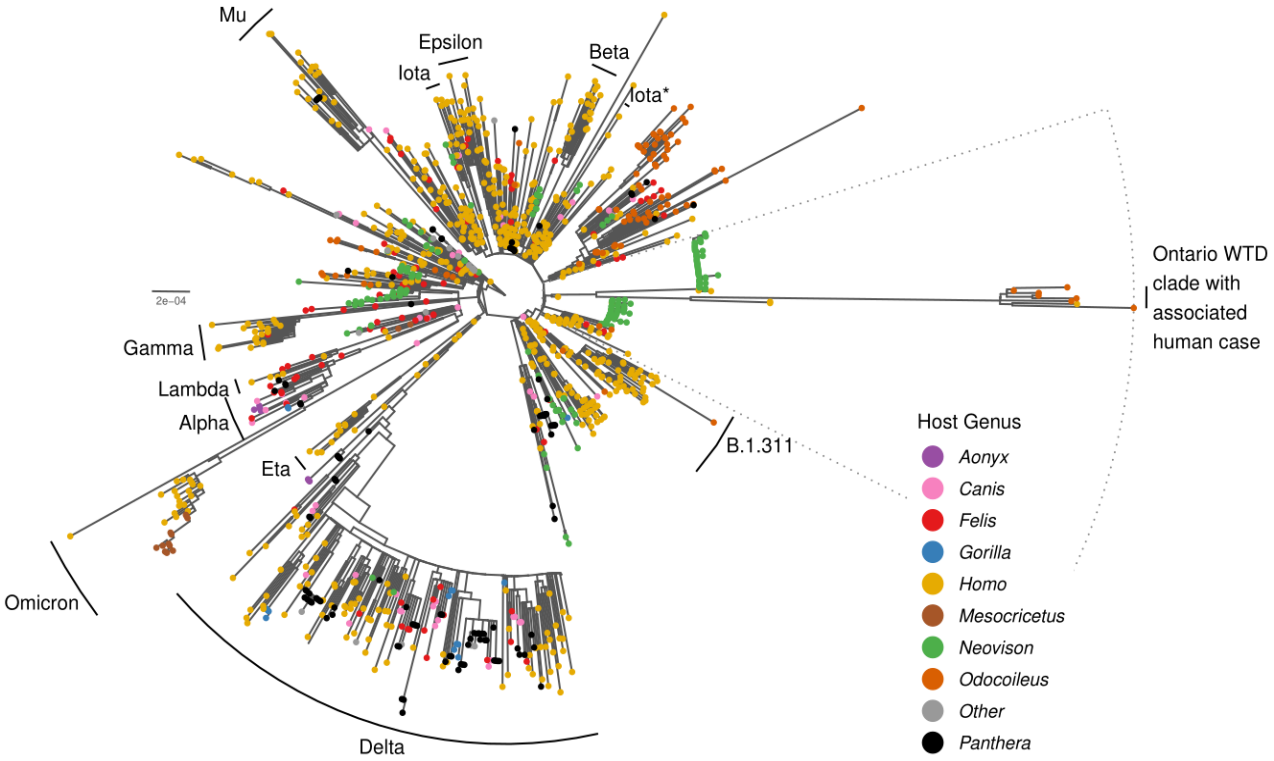

Figure 2: **Maximum-likelihood (ML) phylogeny of WTD-derived viral genomes (and associated human sample) and a representative sample of the global diversity of human and animal-derived SARS CoV-2 (n=3,645). This phylogeny under-samples human sequences to maximise representation of animal-associated diversity. The location of VOCs/VUIs within the tree are annotated and nodes are coloured by host genus (as indicated in the legend). The dotted line indicates the samples selected for the ML analysis shown in Figure 3.**

Given the distorting effects of long-branch attraction and incomplete sampling, there is a degree of uncertainty in the phylogenetic placement of the WTD samples. However, the geographical proximity (Michigan is directly adjacent to the sampling location in Southwestern Ontario) and similar pattern of a mix of animal and human cases (e.g., mink and human cases) provide compelling circumstantial evidence supporting this placement. Given the degree of divergence and potential for phylogenetic biases, we conducted two analyses to examine the possibility of recombination. Using 3Seq (Lam et al., 2018) and bolotie (Varabyou et al., 2021) with datasets representative of human and animal SARS-CoV-2 diversity in GISAID (as of January 2022) there was no indication of recombination in this clade.

Potential deer-to-human transmission

Our phylogenetic analysis also identified a human-derived sequence from Ontario (ON-PHL-21-44225) that was both highly similar (80/90 shared mutations; Table S2) and formed a well-supported monophyletic group (100% UFB) with the WTD samples (Figure 3). The small number of samples and relative diversity within the WTD clade make it difficult to determine the exact relationship between the human sample and other WTD samples (78% UFB for a most recent common ancestor with 4658). However, global (FIgure 2) and focal (Figure 3) ML analyses and an Usher-based (Turakhia et al., 2021) (Figure S6) parsimony analysis all support this human sample belonging to the WTD clade.

The human sequence also has a plausible epidemiological link to the WTD samples as it was collected in the same geographical region (Southwestern Ontario), during the same time period (autumn 2021) after having known close contact with deer. At the time of the human case detection 100% of eligible confirmed PCR positive SARS-CoV-2 samples collected from human cases were requested by Public Health Ontario and Ontario COVID-19 Genomic Network partners for genome sequencing, and no other genetically related human-derived samples were identified. However, it should be noted that not all requested samples are received and/or successfully sequenced, and the surge of Omicron cases necessitated a reduction in the proportion human-derived SARS-CoV-2 sampled for sequencing in the region, moving from 100% to 50%, 20%, and 5% sampling on the 7th, 20th, and 30th of December 2021 respectively (Public Health Ontario, 2022).

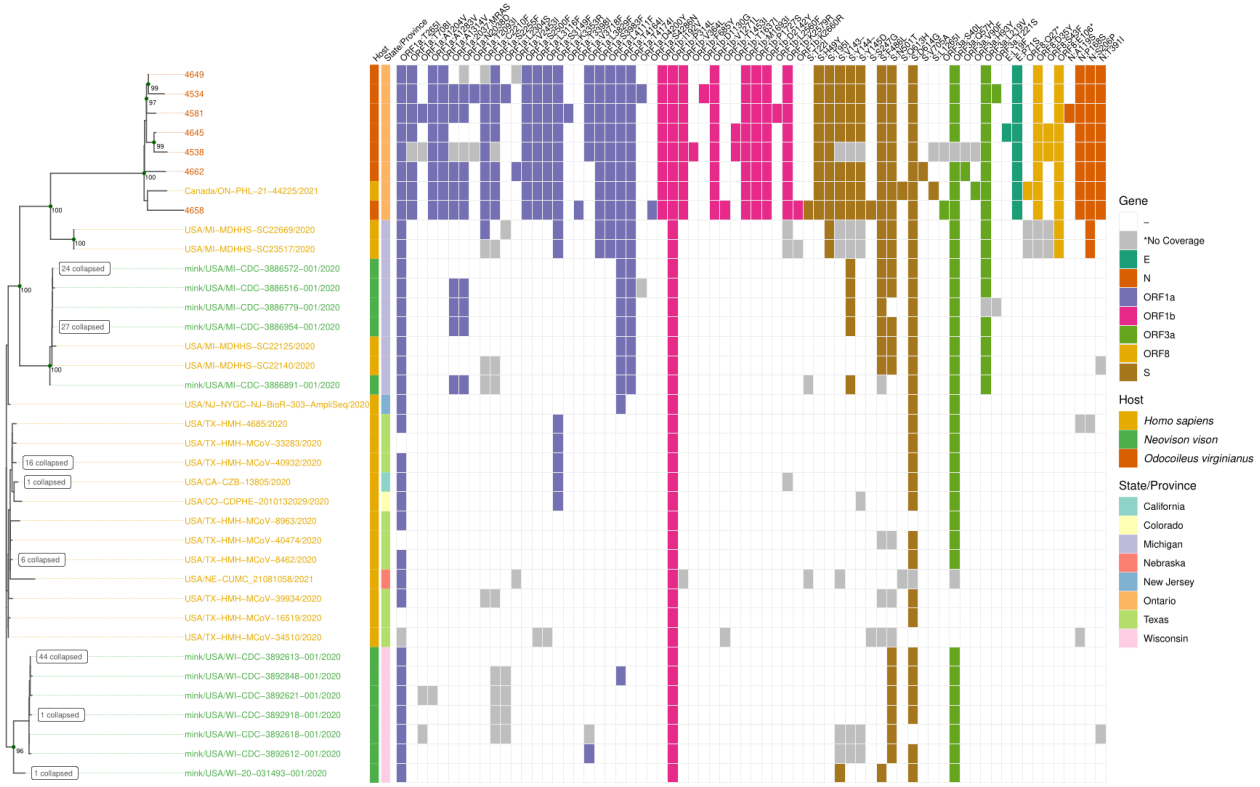

Figure 3: ML phylogeny of 157 genomes selected from global phylogeny (dotted segment in Figure 2) annotated with the presence/absence of amino acid mutations relative to Wuhan Hu-1. Genomes were selected to explore the relationship between Ontario WTD, the related Ontario human sample, and closest B.1 human and mink samples from the USA. Internal nodes in the phylogeny are annotated with

ultrafast bootstrap support values >=95% and leaves with identical amino acid profiles were collapsed as indicated. Host species for each sample is shown by the leaf label colour and first annotation column as per the legend, with geographic location in the second annotation column. Amino acid mutations are coloured by corresponding gene with grey indicating sites that were too poorly covered to determine presence/absence (e.g., sites in partial WTD genomes 4538, 4534).

Zoonosis associated mutations

Using the five high quality completed WTD (4581, 4645, 4649, 4658, and 4662) samples and related human sample (ON-PHL-21-44225), we analysed the prevalence of mutations across GISAID, among VOCs, and from animal-derived samples (the shared non-synonymous subset can be found in Table 1 with full mutation list in Table S2). Of the 76 mutations shared among the 5 high quality WTD and associated human sequences, 51 are in ORF1ab (with 11 and 9 each in Nsp3 and Nsp4, respectively) and 9 are in spike genes. The 6 non-synonymous mutations in S correspond to a 6 nucleotide deletion (V143-Y145), and 5 substitutions (H49Y, T95I, F486L, N501T, D614G). With the exception of the relatively rare but animal associated F486L (439 GISAID sequences as of February 2022), these mutations have been frequently observed in animal-derived viral sequences including bats, cats, hamsters, deer, and mink. Not all spike mutations were conserved across the entire Ontario WTD clade; S:613H and S:L1265I were found only in the human sample and 4 other non-synonymous mutations were found in either 4658 or 4662 WTD samples.

Many non-synonymous mutations had previously been identified in WTD including 16 in at least 3/5 of the WTD samples, S:613H and ORF8:Q27* (associated human sample only), and S:T22I (1/5 ON WTD samples only, but also noted in Delta-like SARS-CoV-2 from deer in Québec (Kotwa et al., 2022)). However, there were also 5 conserved non-synonymous mutations that had not been previously observed in WTD and were relatively rare in GISAID (<1000 sequences as February 2022): ORF1a:insertion2038N/MRASD (n=32, including 31 mink from Michigan, USA), ORF1b:V373L (G14557T, n=425, all human sequences), S:F486L (T23020G, n=439), ORF3a:L219V (T26047G, n=805), and ORF10:L37F (C29666T, n=0).

We compared the ratio of non-synonymous and synonymous mutations (dN/dS) and specific substitution frequencies to other animal genomes, the B.1 lineage in general, and contemporaneous SARS-CoV-2 diversity (Figure 4B). The Ontario WTD (and associated human sequence) had a lower average genome-wide dN/dS (~1) than the other groups. This apparent neutral selection was further supported by the observation that the conserved mutations were distributed proportionally to their associated gene/product length e.g., ORF1ab (54.1 expected vs 51 actual) and spike (9.7 expected vs 9 actual) (Figure S5). Together, these signatures of neutral selection are suggestive of sustained viral transmission with minimal immune pressure in a largely susceptible population of WTD. However, further investigation into the host response and disease course of SARS-CoV-2 in WTD would be required to confirm these inferences.

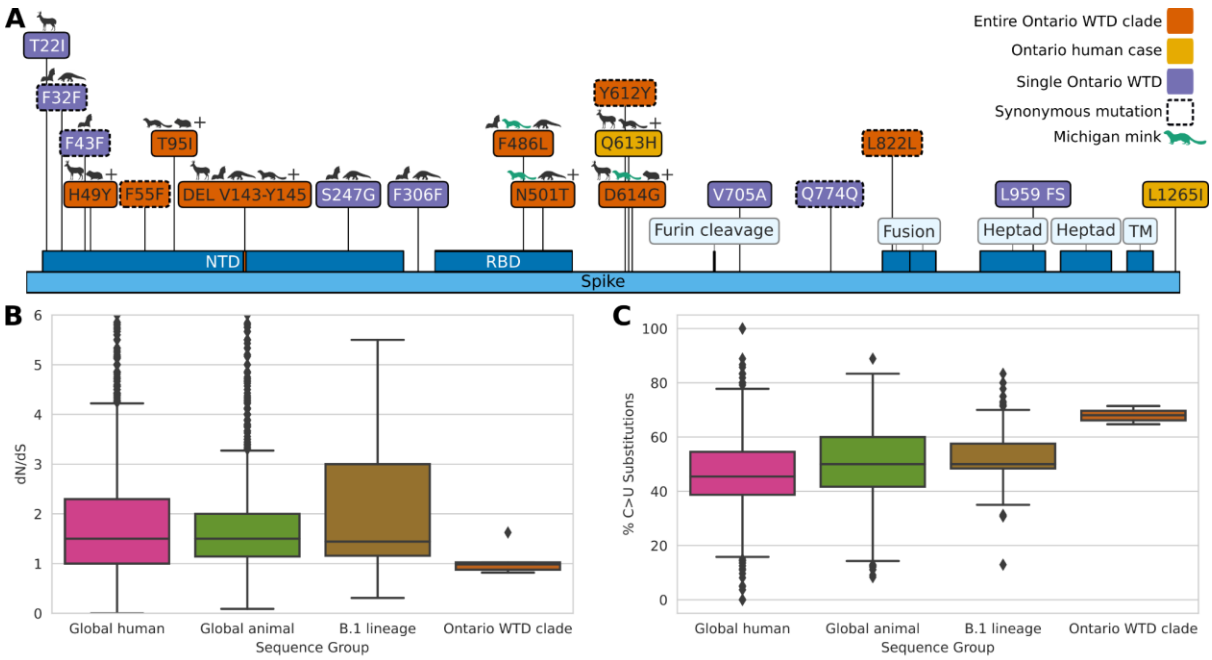

Figure 4: Analysis of variants present in the Ontario WTD clade relative to other animal genomes, the ancestral B.1 lineage, and the global SARS-CoV-2 diversity. (A) Summary of amino acid mutations across the spike protein in the Ontario WTD. Amino acid changes present in all 5 high quality Ontario WTD clade samples (and associated human case) are indicated in orange, those only in the human sample in yellow, and those only in a single WTD genome in purple. Synonymous mutations are indicated by a dashed border. The distribution of mutations across bat, deer, pangolin, and hamster-derived SARS-CoV-2 samples are shown above each mutation by the appropriate symbol. "+" indicates presence of the mutation in additional animal species, and green indicates those in Michigan mink samples. spike annotations were derived from UniProt P0DTC2 (DEL: deletion, FS: frameshift, TM: transmembrane, RBD: Receptor binding domain, NTD: N-terminal domain) and are not shown to scale. (B) Genome-wide dN/dS ratio for a subsample of global human SARS-CoV-2 diversity (earliest and and most recent genomes from each PANGO lineage, n=3,127), global animal diversity (all animal genomes currently in GISAID, n=1,522), B.1 lineage (all genomes assigned to this lineage in GISAID as of January 2022, n=206), and the Ontario WTD clade (5 high quality WTD genomes and associated human case). To ensure consistency between all samples this was calculated directly from nextclade (operating on consensus sequences). (C) Breakdown of % of all consensus substitutions corresponding to a change from a reference C allele to an alternative U allele. The grouping reflects the same sets of sampled diversity as used in (B).

Elevated C>U substitution rate

Changes in the mutational signature of SARS-CoV-2 could be used to trace and understand its spread between human and non-human hosts. An analysis of base substitution frequencies within the Ontario WTD clade (and associated human sample) showed an elevated proportion of mutations involving C>U changes relative to other global, B.1 lineage, and animal-derived viral sequences (Figure 4C, Figure S4). To further explore this, we performed a more robust non-parametric alternative to MANOVA, dbWMANOVA, to determine if the mutational spectra from each host-type differed. The 3,645 samples used to create the global phylogeny were used as input for this analysis. A second analysis, using the 241 samples assigned to Clade 20C (which contains the B.1 lineage), was also conducted. The first analysis included the WTD and human samples from Ontario and Quebec while the second only used the WTD and human samples from Ontario. This revealed a significant difference between the mutational spectrum of samples isolated from

human, deer, and mink hosts (W^*_d = 91.04, $p < 0.001$ and W^*_d = 160.47, $p < 0.001$, respectively) (Table S4, Figure S1). Principal component analysis (PCA) indicated that the majority of this variation (62.2%) corresponded to C>U (PC1) and G>A (PC2) frequencies (Figure S2). Notably, when compared to the recently collected WTD virus samples from Quebec, the location of the Ontario WTD clade (and associated human case) suggest that these samples have been evolving within deer. To understand how much of the variation in the mutational spectra can be ascribed to host-type, a distance-based redundancy analysis using samples from Clade 20C was performed. This analysis found that host-type is an important factor in structuring the mutational spectra of this clade $F(2,238) = 28.12$, $p < 0.001$) (Figure S3).

Preliminary characterization of antigenicity

Considering that mutations in the S gene may lead to immune evasion to antibody responses generated by vaccination, we sought to conduct preliminary analysis of the neutralizing activity of plasma from vaccinated recipients to S proteins identified in this study. Codon-optimized S expression constructs corresponding to the S genes of samples 4581/4645, 4658, and ON-PHL-21-44225 were generated and used, in addition to a construct encoding SARS-CoV-2 S:D614G, to produced lentiviral pseudotypes encoding luciferase. Pseudotypes were then incubated with serial dilutions of sera from naïve individuals or sera collected from vaccinated recipients 28-40 days following a second dose or a third dose of BNT162b2 for 1 hour before infection of 293T-ACE2 cells. Infection was assessed by measuring luciferase activity 72 hours post-infection and neutralization half maximal inhibitory serum dilution (ID50) was determined (Figure 5AB). We found that sera from vaccinated recipients, either after two or three doses, efficiently neutralized all S proteins tested to similar extents. Importantly, we did not observe a difference between the ability of sera to neutralize SARS-CoV-2 D614G or any of the Ontario deer lineage SARS-CoV-2. These results suggest that the mutations in the S proteins tested have no significant antigenic impact.

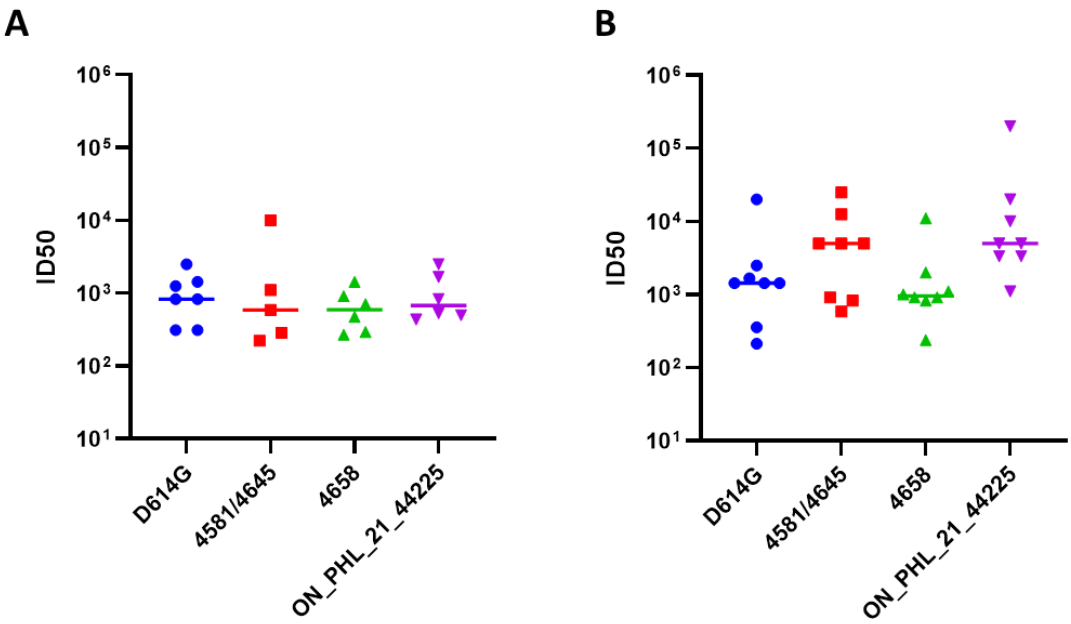

Figure 5. Neutralization of spike proteins by plasma from vaccinated individuals. Lentiviral pseudotypes encoding luciferase and harbouring the indicated spike proteins were incubated with serial dilutions of plasma from vaccinated (2 or 3 doses) or naïve donors individuals as controls for 1 hour at 37°C and then used to infect 293T-ACE2. Infection was measured by quantitating luciferase activity 72 hours post-

infection. Neutralization half maximal inhibitory serum dilution (ID50) for the (A) two or (B) three doses of BNT162b2 sera were determined using a normalized non-linear regression using GraphPad Prism. Significance was determined by analysis of variance (one-way ANOVA) followed by a Dunnett's multiple comparisons test. No significant difference between pseudotypes was observed.

Codon usage comparison to cervid viruses

Analysis of genome composition and codon usage bias (CUB) may provide information on virus evolution and adaptation to host. We assessed whether the codon usage signatures of the Ontario WTD clade is similar to that of other SARS-CoV-2 sequences (samples isolated from Wuhan-Hu-1, Quebec WTD, mink from Canada and the USA), cervid viruses (Epizootic hemorrhagic disease virus (EHDV), Cervid atadenovirus A, elk circovirus), and the *Odocoileus virginianus* (WTD) genome. No apparent differences were observed in codon usage bias between the Ontario WTD clade and other SARS-CoV-2 sequences across the entire coding region of the viral genome. Although some similarity in codon usage bias to Cervid atadenovirus A was observed, generally there were clear differences between SARS-CoV-2 samples and the remaining sequences (Table S5). An analysis of the effective number of codons (ENc), which summarizes the likelihood of using alternative codons (Wright, 1990), showed that SARS-CoV-2 sequences, including those from Ontario had a low codon usage bias. This low codon usage bias could potentially facilitate spillover events involving SARS-CoV-2 since the virus would still be able to replicate efficiently in a new host (Kandeel et al., 2020). A Relative Synonymous Codon Usage (RSCU) analysis for SARS-CoV-2 sequences showed that the majority of the over-represented codons (>1.6) ended with A/U while the majority of the under-represented codons (<0.6) ended with G/C (Hou, 2020). Similar results were seen in our analysis of EHDV and Cervid atadenovirus A. The most over-represented codon in the SARS-CoV-2 was AGA. This codon was also over-represented in Cervid atadenovirus A, elk circovirus, and *Odocoileus virginianus*. The most under-represented codon in SARS-CoV-2, Cervid atadenovirus A, and *Odocoileus virginianus* was UCG. In contrast UUA was over-represented in EHDV, while CUC and ACG was under-represented in EHDV and elk circovirus, respectively.

Discussion

Secondary wildlife reservoirs have the potential to fundamentally alter the ecology of SARS-CoV-2. Following the surveillance of white-tailed deer in Ontario, Canada, we have identified a highly divergent lineage of SARS-CoV-2 with evidence of host adaptation and unsustained deer-to-human transmission. WTD present many attributes important for reservoir sustainability including; social behaviour, high density, highly transient population with significant human-deer interfaces and sylvatic interfaces with other wildlife. A stable reservoir in WTD creates the potential for spillover into human and sylvatic wildlife populations over a broad geographic distance, and determination of this species as a potential secondary reservoir for SARS-CoV-2 merits broader discussion between disease ecologists, virologists, wildlife biologists and public health. Importantly, in contrast to domestic mink which demonstrated the first large-scale inter- and intra- species transmission in animals, onward transmission mitigation within WTD is more challenging than for farmed species.

The limited spatial extent and duration of sampling make it difficult to determine whether this new lineage truly represents a stable secondary wildlife reservoir. However, there are several lines of circumstantial evidence suggestive of a lineage that has been circulating and adapting within a non-human host, including a unique constellation of mutations that has not been previously observed among SARS-CoV-2 lineages. This high degree of divergence (and consequent long branch in the phylogenetic analyses) is indicative of a period of unsampled viral evolution leading to 49 mutations compared to the closest genomes. This is reminiscent of the long branch and viral evolution that led to the Omicron variant, which has recently been linked to a

possible mouse reservoir (Wei et al., 2021). Sustained infection of SARS-CoV-2 within WTD and opportunities for transmission back to humans underscore the potential for WTD to act as an animal reservoir.

Phylogenetic analysis identifies the Ontario WTD lineage as sharing a relatively recent common ancestor with viral sequences from nearby Michigan with a mixed distribution across mink and humans. This includes specific mutational similarities such as a subset of mink sequences in Michigan exhibiting a rare 12-nucleotide insertion in the ORF1a gene that was also present in the Ontario WTD lineage. Two mutations in the S protein, F486L and N501T have been associated with mustelid (mink or ferret) host adaptations, and N501T has been associated with enhanced ACE2 binding and entry into human (Huh7) cells (Han et al., 2021; Lu et al., 2021; Oude Munnink et al., 2021; Starr et al., 2020). Notably, the Ontario WTD SARS-CoV-2 genomes did not harbour the relatively well-described S:Y453F mutation associated with mink and increased replication and morbidity in ferrets, but reduced replication in primary human airway epithelial cells (Zhou et al., 2022). These WTD genomes provide new insights into viral evolution and inferred virus mobility in animal species outside of the human population. It should be noted that many of the mutations found in the divergent Ontario WTD SARS-CoV-2 genomes have either not been described before or are infrequent and uncharacterized.

The mutational spectra of the SARS-CoV-2 genome from WTD, mink, and humans can vary between hosts, as highlighted by significant differences detected within Clade 20C. This is an important observation and provides additional evidence supporting the hypothesis that the mutational spectra can be used to infer viral host-species (Shan et al., 2021; Wei et al., 2021). We found that the frequency of C>U and A>G mutations differs between hosts, which is likely a result of the activity of restriction factors (e.g. the apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like or APOBEC family of mRNA editing proteins), RNA editing enzymes (ADAR1), and reactive oxygen species (De Maio et al., 2021; Mourier et al., 2021; Ringlander et al., 2022; Simmonds & Ansari, 2021). We also observed that the mutational spectrum of the Ontario WTD (and human) sequences are similar to that of other deer. However, this observation does not imply that mutations found in these sequences are the same as those found in other deer. Rather, it is likely that the interaction between the viral genome and various host factors will alter the mutation spectra in broad, host-specific ways (Shan et al., 2021; Wei et al., 2021). When placed in the context of the broader literature, our results provide further evidence this lineage of SARS-CoV-2 likely evolved in deer over time.

It should be noted there were some limitations in genome quality and coverage that may have resulted in failure to detect additional mutations that were present. All Ontario WTD clade samples (including the associated human case) had missing terminal domains and contained internal regions with no or low coverage when sequenced using the ARTIC v4 amplicon scheme. This is a widespread issue that may explain the rarity of the 3' proximal ORF10:L37F in GISAID. Significantly in our samples this meant there was no or <10x coverage in all 5 WTD sequences from ~27000-27177 (dropout of ARTICv4 amplicons 90-91) which includes regions of the M gene. However, by combining the ARTIC v4 sequencing with additional sequencing using probe-based enrichment we were able to compensate for this dropout and generate high coverage and completeness (<100 positions with no coverage in all WTD and <100 positions with <10X coverage in 3/5 WTD genomes; see Table S3).

The neutral non-synonymous to synonymous mutation ratio (dN/dS ~ 1) and evenly distributed mutations in the WTD lineage contrasts with the signatures of strong selection in the equivalently divergent Omicron VOC (dN/dS ~6 and accumulation of spike mutations). In combination with the most recent common ancestor being phylogenetically distant and dating from 2020, we infer that the WTD lineage likely diverged in 2020 and has been maintained in wildlife under relatively neutral selection since that time. It is possible that the absence of pre-existing host immunity has permitted genetic drift to drive accumulation of neutral

mutations (in combination with acquisition of mutations associated with adaptation to animals). Although we cannot be certain whether this is due to a more immune tolerant phenotype in deer, characterization of host response in this species is warranted to determine how deer innate immunity shapes viral adaptation and ecology. The presence of subclinical SARS-CoV-2 related disease in deer also supports this possibility, however substantial gaps in the current knowledge of SARS-CoV-2 pathogenesis in this and other wildlife species persist.

With respect to the initial transmission to deer, it is unclear if spillover occurred directly from humans to deer, or if an intermediate host such as mink or other yet undefined species was involved. The long branch length and period of unsampled evolution provide a number of possible scenarios (Figure 6). It is clear however, that numerous mutations accumulated during this time, including many that appear zoonotic in origin. The human sample branching among the Ontario WTD clade lies within a relatively small number of deer samples, making determining the exact relationship between the human and WTD viruses challenging (78% UFB). This clade could represent a spillover into WTD with a human spillback or the emergence of a virus reservoir in a wildlife species infecting both human and WTD. However, the epidemiological data, paucity of SARS-CoV-2 surveillance in WTD relative to human cases, and potential mink-human spillback in related viruses suggests spillover in WTD followed by deer-to-human transmission is the most likely scenario (Figure 6: Scenario 1).

Rapid characterization of this new lineage from biological and epidemiological perspectives is critical to understanding viral transmission, immune evasion, and disease in both wildlife and humans. Therefore, we performed a preliminary characterization of the antigenicity of the spike proteins identified in this study. Using lentiviral pseudotypes, we found that sera collected from vaccinated individuals after a second or a third dose of BNT162b2 robustly neutralized infection mediated by all spike proteins tested (Figure 5). Neutralization efficiencies were similar to those measured with pseudotypes harbouring SARS-CoV-2 S:D614G suggesting that the mutations in the spike proteins in the Ontario WTD lineage do not have significant impact on spike antigenicity (acknowledging the limited number of plasma tested). More work is needed to determine the potential roles of the mutations on spike functions, and to understand the pathogenesis and transmission phenotypes of this virus.

At this time, there is no evidence of recurrent deer to human or sustained human to human transmission of the Ontario WTD SARS-CoV-2 clade. However, the emergence of Omicron and the end of deer hunting season has meant both human and WTD testing and genomic surveillance in this region has been limited since these samples were collected. Therefore, we cannot determine with certainty whether the lack of additional human cases reflects no onward transmission from the human case, no further spillover events from WTD, or limited genomic surveillance. Enhanced surveillance is of particular importance given human population density and mobility in the region, coupled with WTD population dynamics.

This work underscores the need for a broader international One Health lens to identify new intermediate or reservoir hosts capable of driving sustained transmission and divergent viral evolution. Selective advantage has led to the emergence of new variants outcompeting those in circulation. However, the absence of a variant in the human population does not mean universal absence of variants across a much broader range of under-sampled potential host species. To date, many sampling strategies have been based on access and convenience. Focusing efforts at human-animal interfaces and integration of human epidemiological data would enable analysis of determinants of spillover and inter-species transmission. A broader analysis examining human drivers of spillover and spillback and knock-on effects on wildlife and human health is urgently needed to identify, develop and implement mitigation strategies, beginning with reducing viral activity in humans.

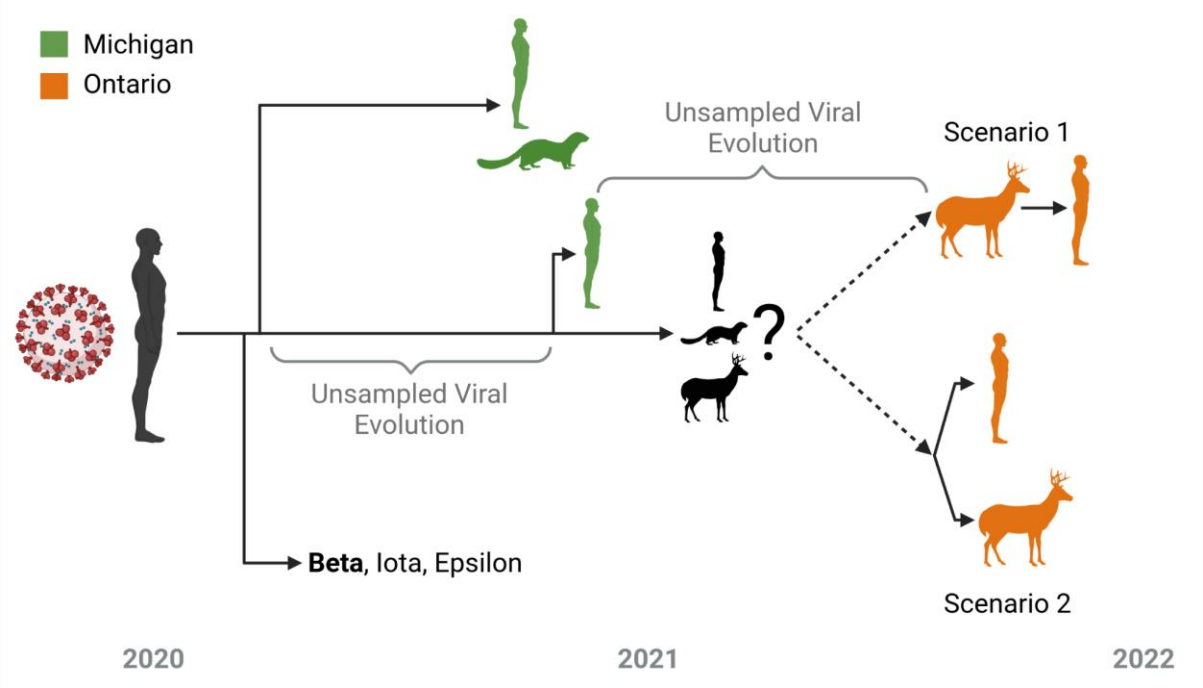

Figure 6: Overview of potential zoonotic scenarios underpinning the evolution of the Ontario WTD clade (and associated human case). The timeline and approximate relationship between the Beta VOC (bold), Iota/Epsilon VUIs, and viral samples in WTD, humans, and mink from both Michigan (green) and Ontario (orange) are displayed. As it likely emerged during one of the indicated poorly sampled periods of viral evolution, it is unclear whether the viral ancestor of the Ontario WTD clade was from an animal (e.g., mink or WTD) or human reservoir. From this ancestor, there was either a spillback transmission from WTD to human (Scenario 1) or the emergence of a virus infecting both human and WTD (Scenario 2).

Table 1: Summary of non-synonymous mutations found across all high quality Ontario WTD clade samples (including associated human case) including their prevalence within GISAID and across animal-derived samples. A full table containing all mutations can be found as Table S2.

Mutation[a] / GISAID Mutation Notation	Total Worldwide (GISAID)[b]	Total Canada (GISAID)[b]	VOC[c]	Non-Human Host (Prevalence)[d]	Previously in deer?
ORF1a:T265I (C1059T) / NSP2_T85I	423,398	14,605	α, δ, ο	dog (12), cat (26), white-tailed deer (92), tiger (21), snow leopard (1), cougar (1), mink (188; 56 MI)	yes
ORF1a:T708I (C2388T) / NSP2_T528I	12,822	226	α, δ, ο	white-tailed deer (1), Chiroptera (4), mink (1; 0 MI)	yes
ORF1a:A1283V (C4113T) / NSP3_A465V	15,787	278	α, δ, ο	cat (1), gorilla (4), mink (3, 0 MI)	no

ORF1a:A1314V (C4206T) / NSP3_A496V	11,484	84	α, δ, ο	Only human	no
ORF1a:N2038delinsMRASD (A6377ATGCGTGCGTCGG) / NSP3_ins1219MRAS[e]	32	1	n/a	mink (31; 31 MI)	no
ORF1a:N2038delinsMRASD (A6377ATGCGTGCGTCGG) / NSP3_N1220D	506	64	α, δ, ο	mink (31; 31 MI)	no
ORF1a:C2210F (G6894T) / NSP3_C1392F[f]	4,609	51	α, δ, ο	mink (2; 0 MI)	no
ORF1a:S2255F (C7029T) / NSP3_S1437F	8,474	11	α, δ, ο	white-tailed deer (3)	yes
ORF1a:S2500F (C7764T) / NSP3_S1682F	19,988	486	α, δ, ο	Chiroptera (1)	no
ORF1a:L3116F (C9611T) / NSP4_L353F	11,563	672	α, δ, ο	Chiroptera (4), white-tailed deer (1)	yes
ORF1a:S3149F (CC9711TT) / NSP4_S386F	3,649	72	α, δ, ο	white-tailed deer (30), Chiroptera (3)	yes
ORF1a:K3353R (A10323G) / NSP5_K90R	155,729	4,411	α, δ, ο	dog (3), cat (4), white-tailed deer(2), mink (1; 0 MI)	yes
ORF1a:L3829F (C11750T) / NSP6_L260F[g]	49,761	4,875	α, δ, ο	hamster (1), white-tailed deer (5), mink (10; 0 MI)	yes
ORF1a:S3983F (C12213T) / NSP8_S41F	2,533	28	α, δ, ο	Only human	no
ORF1a:L4111F (C12596T) / NSP8_L169F	927	48	α, δ, ο	white-tailed deer (1)	yes
ORF1a:T4164I (C12756T) / NSP9_T24I	17,248	462	α, δ, ο	mink (58; 56 MI)	no
ORF1a:T4174I (C12786T) / NSP9_T34I	12,186	791	α, δ, ο	mink (56; 56 MI)	no
ORF1b:I192V (A14041G) / NSP12_I201V	1,922	40	α, δ, ο	Chiroptera (3)	no

ORF1b:P314L (C14408T) / NSP12_P323L	8,192,005	274,376	α, δ, ο	otter, binturong, dog, monkey, hyena, cat, gorilla, hippo, hamster, ferret, ferret, coati, mink (1,163; 56 MI), white-tailed deer, lion, tiger sp., leopard, snow leopard, fishing cat, cougar	yes
ORF1b:V364L (G14557T) / NSP12_V373L	425	3	α, δ, ο	Only human	no
ORF1b:V1271L (G17278T) / NSP13_V348L	9,810	100	α, δ, ο	white-tailed deer (1), snow leopard (4)	yes
ORF1b:M1693I (G18546T) / NSP14_M169I	2,880	49	α, δ, ο	Only human	no
ORF1b:P1727S (C18646T) / NSP14_P203S	4,068	324	α, δ, ο	Only human	no
ORF1b:D2142Y (G19891T) / NSP15_D91Y	7,116	93	α, δ, ο	Only human	no
ORF1b:K2579R (A21203G) / NSP16_K182R	2,106	8	α, δ, ο	Chiroptera (3)	no
S:H49Y (C21707T) / Spike_H49Y	12,853	305	α, δ, ο	cat (1), hamster (4), Roborovski hamster (1), white-tailed deer (3)	yes
S:T95I (C21846T) / Spike_T95I	3,026,465	42,004	α, δ, ο	dog, hyena, cat, gorilla, hippo, tiger, lion, mink (7; 0 MI), hamster	no
S:V143_Y145DELINSD (GTTTATT21989G) / Spike_V143/V145del[h]	1,305,753	20,835	α, δ, ο	cat (1), pangolin (12), white-tailed deer (13), Chiroptera (7), mink (73; 0 MI)	yes
S:F486L (T23020G) / Spike_F486L	439	2	α, δ, ο	Chiroptera (1), mink (295; 55 MI), pangolin (5)	no
S:N501T (A23064C) / Spike_N501T	5,789	29	α, δ, ο	cat (1), mink (223; 54 MI), pangolin (5)	no

S:D614G (A23403G) / Spike_D614G	8,334,179	277,501	α, δ, o	otter, binturong, dog, monkey, hyena, cat, gorilla, hippo, hamster, mouse, ferret, coati, mink (1165; 56MI), white-tailed deer, lion, tiger, snow leopard, leopard, fishing cat, cougar	yes
ORF3a:Q57H (G25563T) / NS3_Q57H	551,554	38,477	α, δ, o	dog (16), cat (35), pangolin (1), ferret (1), tiger (11), white-tailed deer(87), tiger (21), snow leopard (7), cougar(1), mink (267; 56 MI)	yes
ORF3a:L219V (T26047G) / NS3_L219V	805	2	α, δ, o	Chiroptera (5), mink (185; 55 MI)	no
E:P71S (C26455T) / E_P71S	5,401	74	α, δ, o	dog (1), Chiroptera (3)	no
ORF8:D35Y (G27996T) / NS8_D35Y	6,725	55	α, δ, o	Only human	no
ORF8:E106* (G28209T) / NS8_E106stop[i]	35,425	3,127	α, δ, o	dog (1)	no
N:P168S (C28775T) / N_P168S	5,168	443	α, δ, o	Only human	no
N:S206P (T28889C) / N_S206P	1,251	1	α, δ, o	Only human	no
N:T391I (C29445T) / N_T391I	20,642	1,211	α, δ, o	cat (3), lion (1), white-tailed deer (26)	yes
ORF10:L37F (C29666T)	n/a	n/a	n/a	n/a	no

Methods

Deer sample collection and study area

Between November 1 and December 31, 2021, adult and yearling free-ranging WTD were sampled as part of the Ontario Ministry of Northern Development, Mines, Natural Resources and Forestry's (NDMNRF) annual Chronic Wasting Disease (CWD) surveillance program. Samples were collected from hunter-harvested deer in Southwestern and Eastern Ontario and included nasal swabs and retropharyngeal lymph nodes (RPLNs).

All samples were collected by staff who wore a mask and disposable gloves while sampling. Nasal swabs were stored in individual 2mL tubes with ~1mL of universal transport media (UTM; Sunnybrook Research Institute) and RPLN tissues were stored dry in 2 mL tubes. After collection, samples were immediately chilled on ice packs then transferred to a -20^{0}C freezer where they were held for up to one week. Samples were then transferred to a -80^0 C freezer where they were held until analysis. Location and date of harvest and demographic data (age/sex) were recorded for each animal when available.

PCR screening and detection

RNA extractions and PCR testing of samples collected from deer were performed at the Sunnybrook Research Institute (SRI) in Toronto, Ontario. RNA extractions were conducted with 140 µL of nasal swab sample spiked with Armored RNA enterovirus (Asuragen; https://www.asuragen.com) via the Nuclisens EasyMag using Generic Protocol 2.0.1 (bioMérieux Canada Inc., St-Laurent, QC, Canada) according to manufacturer's instructions; RNA was eluted in 50 µL. Tissue samples were thawed, weighed, minced with a scalpel, and homogenized in 600 µL of lysis buffer using the Next Advance Bullet Blender (Next Advance, Troy, NY USA) and a 5mm stainless steel bead at 5 m/s for 3 minutes. RNA from 30 mg tissue samples were extracted using Specific Protocol B 2.0.1 via Nuclisens EasyMag; RNA was eluted in 50 µL. Reverse-transcription polymerase chain reaction (RT-PCR) was performed using the Luna Universal Probe One-Step RT-qPCR kit (New England BioLabs; https:// www.neb.ca). The 5' untranslated region (UTR) and the envelope (E) gene were used for SARS-CoV-2 RNA detection (LeBlanc et al., 2020). Quantstudio 3 software (Thermo Fisher Scientific; https://www.thermofisher.com) was used to determine the cycle threshold (Ct). All samples were run in duplicate and samples with Ct<40 for both gene targets and Armored RNA enterovirus in at least one replicate were considered presumed positive. For tissue samples, the presence of inhibitors was assessed by a 1:5 dilution of one of the replicates. Samples were considered inconclusive if no Armored enterovirus was detected or if only one gene target was detected and were re-extracted for additional analysis. Samples were considered indeterminate if inconclusive after re-extraction or if no original material was left. Presumed positive samples were further analysed for human RNAse P to rule out potential human contamination (Lu et al., 2020).

Original material from presumed positive samples detected at SRI were sent to the Canadian Food Inspection Agency (CFIA) for confirmatory PCR testing. The MagMax CORE Nucleic Acid Purification Kit (ThermoFisher Scientific) and the automated KingFisher Duo Prime magnetic extraction system was used to extract total RNA spiked with Armored RNA enterovirus. The enteroviral armored RNA was used as an exogenous extraction control. The E and nucleocapsid (N) genes were used for confirmatory SARS-CoV-2 RNA detection (Lu et al., 2020). Master mix for qRT-PCR was prepared using TaqMan Fast Virus 1-step Master Mix (ThermoFisher Scientific) according to manufacturer's instructions. Reaction conditions were 50°C for 5 minutes, 95°C for 20 seconds, and 40 cycles of 95°C for 3 seconds then 60°C for 30 seconds. Runs were performed by using a 7500 Fast Real-Time PCR System (Thermofisher, ABI). Samples with Ct <36 for both gene targets were considered positive.

WGS sequencing

WGS was performed at both SRI and CFIA using independent extractions and sequencing methods. At SRI, DNA was synthesised from extracted RNA using 4 µL LunaScript RT SuperMix 5X (New England Biolabs, NEB, USA) and 8 µL nuclease free water, were added to 8 µL extracted RNA. cDNA synthesis was performed under the following conditions: 25 °C for 2 min, 55 °C for 20 min, 95 °C for 1 min, and holding at 4 °C.

The ARTIC V4 primer pool (https://github.com/artic-network/artic-ncov2019) was used to generate amplicons from the cDNA. Specifically, two multiplex PCR tiling reactions were prepared by combining 2.5 µL cDNA with 12.5 µL Q5 High-Fidelity 2X Master Mix (NEB, USA), 6µL nuclease free water, and 4 µL of

respective 10 µM ARTIC V4 primer pool (Integrated DNA Technologies). PCR cycling was then performed in the following manner: 98 °C for 30 s followed by 35 cycles of 98 °C for 15 s and 63 °C for 5 min.

PCR reactions were then both combined and cleaned using 1X ratio Sample Purification Beads (Illumina) at a 1:1 bead to sample ratio. The amplicons were quantified using the Qubit 4.0 fluorometer using the 1X dsDNA HS Assay Kit (Thermo Fisher Scientific, USA) and sequencing libraries prepared using the Nextera DNA Flex Prep kit (Illumina, USA) as per manufacturer's instructions. Paired-end (2×150 bp) sequencing was performed on a MiniSeq with a 300–cycle reagent kit (Illumina, USA) with a negative control library with no input SARS-CoV-2 RNA extract.

WGS performed at CFIA used extracted nucleic acid quantified using the Qubit™ RNA High Sensitivity (HS) Assay Kit on a Qubit™ Flex Fluorometer (Thermo Fisher Scientific). 11uL or 200ng of total RNA was subject to DNase treatment using the ezDNase™ Enzyme (Thermo Fisher Scientific) according to the manufacturer's instructions. DNase treated RNA was then used for library preparation and target sequence capture according to the ONETest™ Coronaviruses Plus Assay protocol (Fusion Genomics (Zhan et al., 2021)). The enriched libraries were then quantified using the Qubit™ 1x dsDNA HS Assay Kit on a Qubit™ Flex Fluorometer (Thermo Fisher Scientific) and subsequently pooled in equimolar amounts prior to fragment analysis on 4200 TapeStation System using the D5000 ScreenTape Assay (Agilent). The final pooled library was sequenced an Illumina MiSeq using a V3 flowcell and 600 cycle kit (Illumina).

Human specimens are received at Public Health Ontario Laboratory (PHOL) for routine SARS-CoV-2 diagnostic testing (RT-PCR) from multiple healthcare settings across the province, including hospitals, clinics and COVID-19 assessment centres. The human sample (ON-PHL-21-44225) was sequenced at PHOL using an Illumina-based ARTIC V4 protocol (10.17504/protocols.io.b5ftq3nn), similar to the deer sequencing methods. Briefly, cDNA was synthesized using LunaScript reverse transcriptase (New England BioLabs). Amplicons were generated with premixed ARTIC V4 primer pools (Integrated DNA Technologies). Amplicons from the two pools were combined, purified with AMPure XP beads (Beckman Coulter) and quantified. Genomic libraries were prepared using the Nextera XT DNA Library Preparation Kit (Illumina, San Diego, CA) and genomes were sequenced as paired-end (2 × 150-bp) reads on an Illumina MiSeq instrument.

Genomic analysis

Paired-end illumina reads from ARTIC V4 and Fusion Genomics sequencing were initially analysed separately with the nf-core/viralrecon Nextflow workflow (v2.3) (Di Tommaso et al., 2017; Ewels et al., 2020; Patel et al., 2022) which ran: FASTQC (v0.11.9) (Andrews, 2010) read-level quality control, fastp (v0.20.1) (Chen et al., 2018) quality filtering and adapter trimming, Bowtie2 (v2.4.2) (Langmead & Salzberg, 2012) read mapping to Wuhan-Hu-1 (MN908947.3) (Wu et al., 2020) SARS-CoV-2 reference, Mosdepth (v0.3.1)(Pedersen & Quinlan, 2018)/Samtools (v.1.12) (Li et al., 2009) read mapping statistics calculation, iVar (v1.3.1) (Grubaugh et al., 2019) ARTIC V4 primer trimming, variant calling, and consensus generation; SnpEff (v5.0) (Cingolani, Platts, et al., 2012)/SnpSift (v4.3t) (Cingolani, Patel, et al., 2012) for variant effect prediction and annotation; and Pangolin (v3.1.20) (O'Toole et al., 2021) with PangoLEARN (2022-01-05), Scorpio (v0.3.16) (Colquhoun & Jackson, 2021), and Constellations (v.0.1.1) was used for PANGO lineage (Rambaut et al., 2020) assignment. iVar primer trimmed soft-clipped read alignments were converted to hard-clipped alignments with fgbio ClipBam (http://fulcrumgenomics.github.io/fgbio/). Reads from hard-clipped BAM files were output to FASTQ files with "samtools fastq". nf-core/viralrecon was re-run in amplicon mode without iVar primer trimming on the combined Fusion Genomics and ARTIC V4 primer trimmed FASTQ reads to generate the variant calling, variant effect and consensus sequence results used in downstream analyses. Additional quality control steps to check for negative control validity, drop-out, sample cross-contamination, and excess ambiguity were performed using ncov-tools v1.8.0 (Jared Simpson & de Borja, 2020). The mutations

identified by the Nextclade (v1.10.2) (Aksamentov et al., 2021) with 2022-01-05 database and xlavir (v0.6.1) report were manually searched in outbreak.info's "Lineage | Mutation Tracker" (on 2022-02-02) (Tsueng et al., 2022) to get information on the prevalence of observed mutations globally and within Canada. Mutations were also investigated for presence in specific lineages including VOCs, Michigan mink samples, and other animal samples. Finally, mutations were searched in GISAID (on 2022-02-02) to tally the number of non-human hosts each mutation had been observed in.

Phylogenetics

To evaluate possible sampling biases due to the poorly defined and diverse B.1 and B.1.311 lineages and select closely related publicly available sequences for further phylogenetic analysis, an UShER (https://genome.ucsc.edu/cgi-bin/hgPhyloPlace) (Turakhia et al., 2021) based phylogenetic placement analysis was performed using the 7,217,299 sample tree (derived from UShER placement of GISAID, GenBank, COG-UK, and CNCB onto 13-11-20 sarscov2phylo ML tree) via the SHUShER web-portal (shusher.gi.ucsc.edu). Phylogenetic analyses were performed using CFIA-NCFAD/scovtree Nextflow workflow (v1.6.0) (https://github.com/CFIA-NCFAD/scovtree/) with the consensus sequences contextualised with closely related sequences identified by UShER and randomly sampled representative sequences from major WHO SARS-CoV-2 clades from GISAID (Shu & McCauley, 2017) (downloaded 2022-02-10). This workflow generated a multiple sequence alignment using Nextalign CLI (v1.10.1) (Aksamentov et al., 2021) and inferred a maximum-likelihood (ML) phylogeny using IQ-TREE (v2.2.0_beta) (Minh et al., 2020) using the GTR model for visualisation with Phylocanvas (Abudahab et al., 2021) via shiptv (v0.4.1) (https://github.com/CFIA-NCFAD/shiptv) and ggtree (Yu et al., 2017).

A subset of 157 taxa from an ancestral clade of the Ontario WTD and human clade were selected from the global phylogenetic tree shown in Figure 2 to generate the phylogenetic tree shown in Figure 3. Multiple sequence alignment of this subset of sequences was performed with MAFFT (v7.490) (Katoh & Standley, 2013). A maximum-likelihood phylogenetic tree was inferred with IQ-TREE (v2.2.0_beta) using the GTR model and 1000 ultrafast bootstrap replicates (Hoang et al., 2018). Nextclade (v1.10.2) analysis was used to determine amino acid mutations and missing or low/no coverage regions from the sample genome sequences. Amino acid mutation profiles were determined relative to the Ontario WTD and human samples, discarding mutations that were not present in any of the Ontario samples. Taxa with duplicated amino acid mutation profiles were pruned from the tree, keeping only the first observed taxa with a duplicated profile.

Recombination analyses were performed using 3Seq (v1.7) (Lam et al., 2018) and Bolotie (e039c01) (Varabyou et al., 2021)**.** Specifically, 3Seq was executed with WTD+Human sequences and the most recent example of each lineage found in Canada and closest samples in GISAID in subtree (n=595). Bolotie was executed using the WTD+Human sequences and two datasets, the provided pre-computed 87,695 probability matrix and a subsample of of the earliest and latest example of each lineage in GISAID with all animal-derived samples and closest usher samples (n=4,688). Sequence statistics (dN/dS and C>T) were directly calculated from nextclade results (v1.10.2 with 2022-01-05 database). Additional figures were generated and annotated using BioRender (BioRender, 2022) and Inkscape (Inkscape Project, 2020).

Analysis of mutational spectrum

The mutational spectra was created using a subset of 3,645 sequences used to create the high-quality global phylogeny. Included in this dataset are a collection of seven unique WTD samples from Ontario (samples 4538, 4534, 4662, 4649, 4581, 4645, and 4658; the 5 high quality genomes plus 2 genomes with lower coverage), three WTD samples from Quebec (samples 4055, 4022, 4249), and one human sample from Ontario (ON-PHL-21-44225). The counts for each type of nucleotide change, with respect to the reference strain, were used to create a 12-dimensional vector. The final dataset consisted of all human, mink, and deer

samples originating from the United States of America and Canada with at least 15 mutations. The counts were converted into the mutation spectrum by simply dividing each count by the sum of the counts in each sample(Wei et al., 2021). To investigate differences in mutation spectra between hosts, a distance-based Welch MANOVA (dbWMANOVA) was run (Hamidi et al., 2019). If a significant difference was detected, a pairwise distance-based Welch t-Test was used to identify which pair of hosts significantly differed (Alekseyenko, 2016). These tests were used since they are more robust on unbalanced and heteroscedastic data (Alekseyenko, 2016; Hamidi et al., 2019). We also used a distance-based redundancy analysis (dbRDA), performed using the 'vegan' package in R v4.1.2, to analyze the effect of host in Clade 20C (Legendre & Anderson, 1999). This analysis decomposes the composition of the mutational spectra into two components: that which can be explained by a set of constraining variables (in this case the host type) and that which remains unexplained. We used this test to investigate how much of the variation in the mutational spectra can be accounted for by the host type. Type II permutation tests using the 'RVAideMemoire' package were conducted in R v4.1.2 using the fitted model to determine the significance of the model and the constraining variables (Hervé, 2021; Orians et al., 2019).

Codon-optimized spike constructs, pseudotype production and neutralization assays

Expression constructs of S mutants corresponding to samples 4581/4645 (S:H49Y, S:T95I, S:Δ143-145InsD, S:F486L, S:N501T, S:D614G), 4658 (S:T22I, S:T95I, S:Δ143-145InsD, S:S247G, S:F486L, S:N501T, S:D614G) and ON-PHL-21-44225 (S:H49Y, S:T95I, S:Δ143-145InsD, S:F486L, S:N501T, S:Q613H, S:D614G) were generated by overlapping PCR as described previously (Chatterjee et al., 2022). All constructs were cloned in pCAGGS and verified by Sanger sequencing.

HEK293T cells (ATCC) were cultured in Dulbecco's Minimum Essential Medium (DMEM) supplemented with 10% fetal bovine serum (FBS, Sigma), 100 U/mL penicillin, 100 µg/mL streptomycin, and 0.3 mg/mL L-glutamine (Invitrogen) and maintained at 37°C, 5% CO_2 and 100% relative humidity. HEK293T seeded in 10-cm dishes were co-transfected with lentiviral packaging plasmid psPAX2 (gift from Didier Trono, addgene #12260), lentiviral vector pLentipuro3/TO/V5-GW/EGFP-Firefly Luciferase (gift from Ethan Abela, addgene#119816), and plasmid encoding the indicated S construct at a 5:5:1 ratio using jetPRIME transfection reagent according to the manufacturer protocol. Twenty-four hours post-transfection, media were changed and supernatants containing lentiviral pseudotypes were harvested 48 h post-transfection, filtered with a 0.45 µM filter and stored at -80°C until use.

Blood samples were obtained from donors at Mount Sinai Hospital who consented to participate in this study (REB 22-0030-E). Plasma of SARS-CoV-2 naïve and naïve-vaccinated (28-40 days post two- or three- doses of BNT162b2) donors were collected, heat-inactivated for 1 h at 56°C, aliquoted and stored at -80°C until use.

HEK293T stably expressing human ACE2 (293T-ACE2, kind gift of Hyeryun Choe, Scripps Research) were seeded in poly-D-lysine-coated 96-well plates. The next day, supernatants containing lentiviral pseudotypes were incubated with sera (serially diluted) for 1 hour at 37° and then added to cells in the presence of 5 µg/mL polybrene. Seventy-two hours later, media were removed, and cells were rinsed in phosphate-buffered saline and lysed by the addition of 40 µl passive lysis buffer (Promega) followed by one freeze-thaw cycle. A Synergy Neo2 Multi-Mode plate reader (BioTek) was used to measure the luciferase activity of each well after the addition of 50-100 µl of reconstituted luciferase assay buffer (Promega) as per the manufacturer's protocol. Neutralization half-maximal inhibitory dilution (ID50) was calculated using Graphpad Prism and represents the plasma dilution that inhibits 50% of pseudotype transduction in 293T-ACE2.

Codon usage analysis

Consensus sequences of SARS-CoV-2 samples from this and previous studies and additional sequences gathered from public databases were used. The sequences include the reference SARS-CoV-2 Wuhan-Hu-1 (NCBI NC_045512), SARS-CoV-2 mink/Canada/D/2020 (GISAID EPI_ISL_717717), SARS-CoV-2 mink/USA/MI-20-028629-004/2020 (GISAID EPI_ISL_2834697), *Cervid atadenovirus A* 20-5608 (NCBI OM470968) (Lung et al., 2022), EHDV serotype 2 / strain Alberta (NCBI AM744997 - AM745006), Epizootic Hemorrhagic Disease Virus, EHDV serotype 1 / New Jersey (NCBI NC_013396 - NC_013405), EHDV 6 isolate OV208 (NCBI MG886400 - MG886409), and Elk circovirus Banff/2019 (NCBI MN585201) (Fisher et al., 2020) were imported into Geneious (v.9.1.8) (Kearse et al., 2012). Annotations for the coding sequences of SARS-CoV-2 samples were transferred from the reference sequence SARS-CoV-2 Wuhan-Hu-1 (NC_ 045512) using the Annotate from Database tool. The coding sequences were extracted using the Extract Annotations tool for all viral sequences. An annotated file of the coding sequences for the *Odocoileus virginianus texanus* isolate animal Pink-7 (GCF_002102435.1) genome was downloaded from NCBI (https://ftp.ncbi.nlm.nih.gov/genomes/all/annotation_releases/9880/100/GCF_002102435.1_Ovir.te_1.0/). Coding sequences were input into CodonW (http://codonw.sourceforge.net/) with settings set to concatenate genes and output to file set to codon usage. Codon usage indices were set to the effective number of codons (ENc), GC content of gene (G+C), GC of silent 3rd codon position (GC3s), silent base composition, number of synonymous codons (L_sym), total number of amino acids (L_aa), hydrophobicity of protein (Hydro), and aromaticity of protein (Aromo).

Acknowledgments

We acknowledge the contributions of the Virus Detection, Molecular Diagnostics and DNA Core sections of Public Health Ontario, and in particular Sarah Teatero for leading the genome sequencing efforts. We gratefully acknowledge contributions of SARS-CoV-2 genome sequences from other laboratories through GISAID (Table S6). We wish to thank the licensed Ontario deer hunters who submitted samples for wildlife disease surveillance, the staff of NDMNRF's CWD surveillance program for their assistance with sample collection, and Sarah Hagey for GIS support. We also acknowledge the contributions of CFIA NCFAD's Genomics Unit for their assistance with additional laboratory support and sequencing. S.M. and B.P. are members of the Canada and the Canadian Institutes for Health Research *Coronavirus Variants Rapid Response Network (CoVaRR-Net)*.

Funding: Funding to S.M. was provided by the Public Health Agency of Canada and the Canadian Institutes for Health Research Operating grant: Emerging COVID-19 Research Gaps and Priorities #466984. J.K. was supported by the Association of Medical Microbiology and Infectious Diseases (AMMI) Canada 2020 AMMI Canada/Biomérieux Post Residency Fellowship in Microbial Diagnostics (unrestricted). Funding and computing resources for F.M were provided by the Shared Hospital Laboratory, Dalhousie University, and the Donald Hill Family. Funding for B.P was provided by the Canadian Food Inspection Agency and the Canadian Safety and Security Program. Funding to J.B., T.B., and L.N. was provided by NDMNRF and the Public Health Agency of Canada. Funding for O.L. was from the Canadian Safety and Security Program, Laboratories Canada and CFIA. M.C. is a Canada Research Chair in Molecular Virology and Antiviral Therapeutics (950-232840) and a recipient of an Ontario Ministry of Research, Innovation and Science Early Researcher Award (ER18-14-09).Funding to M.C. was provided by a COVID-19 Rapid Research grant from the Canadian Institutes for Health Research (CIHR, OV3 170632), CIHR stream 1 for SARS-CoV-2 Variant Research. Work from M.C. is also supported by the Sentinelle COVID Quebec network led by the Laboratoire de Santé Publique du Québec (LSPQ) in collaboration with Fonds de Recherche du Québec-Santé (FRQS) and Genome Canada – Génome Québec, and by the Ministère de la Santé et des Services Sociaux (MSSS) and the Ministère de l'Économie et Innovation (MEI).

Author Contribution Statement

Using the Contributor Roles Taxonomy - Conceptualisation: B.P., O.L., J.B., S.M., T.B; Methodology: F.M., O.L., J.K., B.P., D.S.; Software: F.M., P.K., O.L., J.S.; Formal analysis: F.M., P.K., J.G., O.L., J.S.; Investigation: M.C., K.N., P.A., J.B-S., H-Y.C, E.C., W.Y., M.G., M.S., M.P., G.S., A.M-A.; Data Curation: J.K., P.K., L.N., H.M.; Resources: A.M., E.A.; Writing - Original Draft: F.M., J.B., O.L., S.M., B.P., J.G.; Writing - Review & Editing: All authors; Visualisation: F.M., P.K., J.S.; Supervision: J.B., S.M., O.L.; Funding: B.P., S.M., J.K., O.L., J.B., T.B., L.N.

References

Abdel-Moneim, A. S., & Abdelwhab, E. M. (2020). Evidence for SARS-CoV-2 Infection of Animal Hosts. *Pathogens*, *9*(7), E529. https://doi.org/10.3390/pathogens9070529

Abudahab, K., Underwood, A., Taylor, B., Yeats, C., & Aanensen, D. M. (2021). *Phylocanvas.gl: A WebGL-powered JavaScript library for large tree visualisation*. OSF Preprints. https://doi.org/10.31219/osf.io/nfv6m

Aksamentov, I., Roemer, C., Hodcroft, E., & Neher, R. (2021). Nextclade: Clade assignment, mutation calling and quality control for viral genomes. *Journal of Open Source Software*, *6*(67), 3773. https://doi.org/10.21105/joss.03773

Alekseyenko, A. V. (2016). Multivariate Welch t-test on distances. *Bioinformatics*, *32*(23), 3552–3558. https://doi.org/10.1093/bioinformatics/btw524

Andersen, K. G., Rambaut, A., Lipkin, W. I., Holmes, E. C., & Garry, R. F. (2020). The proximal origin of SARS-CoV-2. *Nature Medicine*, *26*(4), 450–452. https://doi.org/10.1038/s41591-020-0820-9

Andrews, S. (2010). *FastQC: a quality control tool for high throughput sequence data.* http://www.bioinformatics.babraham.ac.uk/projects/fastqc

BioRender. (2022). *BioRender*. BioRender.com

Boni, M. F., Lemey, P., Jiang, X., Lam, T. T.-Y., Perry, B. W., Castoe, T. A., Rambaut, A., & Robertson, D. L. (2020). Evolutionary origins of the SARS-CoV-2 sarbecovirus lineage responsible for the COVID-19 pandemic. *Nature Microbiology*, *5*(11), 1408–1417. https://doi.org/10.1038/s41564-020-0771-4

Chandler, J. C., Bevins, S. N., Ellis, J. W., Linder, T. J., Tell, R. M., Jenkins-Moore, M., Root, J. J., Lenoch, J. B., Robbe-Austerman, S., DeLiberto, T. J., Gidlewski, T., Torchetti, M. K., & Shriner, S. A. (2021). SARS-CoV-2 exposure in wild white-tailed deer (Odocoileus virginianus). *Proceedings of the National Academy of Sciences*, *118*(47). https://doi.org/10.1073/pnas.2114828118

Chatterjee, D., Tauzin, A., Laumaea, A., Gong, S. Y., Bo, Y., Guilbault, A., Goyette, G., Bourassa, C., Gendron-Lepage, G., Medjahed, H., Richard, J., Moreira, S., Côté, M., & Finzi, A. (2022). Antigenicity of the Mu (B.1.621) and A.2.5 SARS-CoV-2 Spikes. *Viruses*, *14*(1), 144. https://doi.org/10.3390/v14010144

Chen, S., Zhou, Y., Chen, Y., & Gu, J. (2018). fastp: An ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics*, *34*(17), i884–i890. https://doi.org/10.1093/bioinformatics/bty560

Cingolani, P., Patel, V. M., Coon, M., Nguyen, T., Land, S. J., Ruden, D. M., & Lu, X. (2012). Using Drosophila melanogaster as a Model for Genotoxic Chemical Mutational Studies with a New Program, SnpSift. *Frontiers in Genetics*, *3*, 35. https://doi.org/10.3389/fgene.2012.00035

Cingolani, P., Platts, A., Wang, L. L., Coon, M., Nguyen, T., Wang, L., Land, S. J., Lu, X., & Ruden, D. M. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. *Fly*, *6*(2), 80–92. https://doi.org/10.4161/fly.19695

Colquhoun, R., & Jackson, B. (2021). *Scorpio*. https://github.com/cov-lineages/scorpio

Cui, J., Li, F., & Shi, Z.-L. (2019). Origin and evolution of pathogenic coronaviruses. *Nature Reviews Microbiology*, *17*(3), 181–192. https://doi.org/10.1038/s41579-018-0118-9

Damas, J., Hughes, G. M., Keough, K. C., Painter, C. A., Persky, N. S., Corbo, M., Hiller, M., Koepfli, K.-P., Pfenning, A. R., Zhao, H., Genereux, D. P., Swofford, R., Pollard, K. S., Ryder, O. A., Nweeia, M. T., Lindblad-Toh, K., Teeling, E. C., Karlsson, E. K., & Lewin, H. A. (2020). Broad host range of SARS-CoV-2 predicted by comparative and structural analysis of ACE2 in vertebrates. *Proceedings of the National Academy of Sciences*, *117*(36), 22311–22322. https://doi.org/10.1073/pnas.2010146117

De Maio, N., Walker, C. R., Turakhia, Y., Lanfear, R., Corbett-Detig, R., & Goldman, N. (2021). Mutation Rates and Selection on Synonymous Mutations in SARS-CoV-2. *Genome Biology and Evolution*, *13*(5), evab087. https://doi.org/10.1093/gbe/evab087

Di Tommaso, P., Chatzou, M., Floden, E. W., Barja, P. P., Palumbo, E., & Notredame, C. (2017). Nextflow enables reproducible computational workflows. *Nature Biotechnology*, *35*(4), 316–319. https://doi.org/10.1038/nbt.3820

Ewels, P. A., Peltzer, A., Fillinger, S., Patel, H., Alneberg, J., Wilm, A., Garcia, M. U., Di Tommaso, P., & Nahnsen, S. (2020). The nf-core framework for community-curated bioinformatics pipelines. *Nature Biotechnology*, *38*(3), 276–278. https://doi.org/10.1038/s41587-020-0439-x

Fisher, M., Harrison, T. M. R., Nebroski, M., Kruczkiewicz, P., Rothenburger, J. L., Ambagala, A., Macbeth, B., & Lung, O. (2020). Discovery and comparative genomic analysis of elk circovirus (ElkCV), a novel circovirus species and the first reported from a cervid host. *Scientific Reports*, *10*(1), 19548. https://doi.org/10.1038/s41598-020-75577-6

Frutos, R., & Devaux, C. A. (2020). Mass culling of minks to protect the COVID-19 vaccines: Is it rational? *New Microbes and New Infections*, *38*, 100816. https://doi.org/10.1016/j.nmni.2020.100816

Grubaugh, N. D., Gangavarapu, K., Quick, J., Matteson, N. L., De Jesus, J. G., Main, B. J., Tan, A. L., Paul, L. M., Brackney, D. E., Grewal, S., Gurfield, N., Van Rompay, K. K. A., Isern, S., Michael, S. F., Coffey, L. L., Loman, N. J., & Andersen, K. G. (2019). An amplicon-based sequencing framework for accurately measuring intrahost virus diversity using PrimalSeq and iVar. *Genome Biology*, *20*(1), 8. https://doi.org/10.1186/s13059-018-1618-7

Haagmans, B. L., Al Dhahiry, S. H. S., Reusken, C. B. E. M., Raj, V. S., Galiano, M., Myers, R., Godeke, G.-J., Jonges, M., Farag, E., Diab, A., Ghobashy, H., Alhajri, F., Al-Thani, M., Al-Marri, S. A., Al Romaihi, H. E., Al Khal, A., Bermingham, A., Osterhaus, A. D. M. E., AlHajri, M. M., & Koopmans, M. P. G. (2014). Middle East respiratory syndrome coronavirus in dromedary camels: An outbreak investigation. *The Lancet. Infectious Diseases*, *14*(2), 140–145. https://doi.org/10.1016/S1473-3099(13)70690-X

Hale, V. L., Dennis, P. M., McBride, D. S., Nolting, J. M., Madden, C., Huey, D., Ehrlich, M., Grieser, J., Winston, J., Lombardi, D., Gibson, S., Saif, L., Killian, M. L., Lantz, K., Tell, R., Torchetti, M., Robbe-Austerman, S., Nelson, M. I., Faith, S. A., & Bowman, A. S. (2021). SARS-CoV-2 infection in free-ranging white-tailed deer. *Nature*, 1–8. https://doi.org/10.1038/s41586-021-04353-x

Hallmaier-Wacker, L. K., Munster, V. J., & Knauf, S. (2017). Disease reservoirs: From conceptual frameworks to applicable criteria. *Emerging Microbes & Infections*, *6*(1), 1–5. https://doi.org/10.1038/emi.2017.65

Hamidi, B., Wallace, K., Vasu, C., & Alekseyenko, A. V. (2019). W_{d}^{*}-test: Robust distance-based multivariate analysis of variance. *Microbiome*, *7*(1), 51. https://doi.org/10.1186/s40168-019-0659-9

Han, P., Su, C., Zhang, Y., Bai, C., Zheng, A., Qiao, C., Wang, Q., Niu, S., Chen, Q., Zhang, Y., Li, W., Liao, H., Li, J., Zhang, Z., Cho, H., Yang, M., Rong, X., Hu, Y., Huang, N., … Qi, J. (2021). Molecular insights into receptor binding of recent emerging SARS-CoV-2 variants. *Nature Communications*, *12*(1), 6103. https://doi.org/10.1038/s41467-021-26401-w

Hervé, M. (2021). *RVAideMemoire: Testing and Plotting Procedures for Biostatistics.*

Hoang, D. T., Chernomor, O., von Haeseler, A., Minh, B. Q., & Vinh, L. S. (2018). UFBoot2: Improving the Ultrafast Bootstrap Approximation. *Molecular Biology and Evolution*, *35*(2), 518–522. https://doi.org/10.1093/molbev/msx281

Hou, W. (2020). Characterization of codon usage pattern in SARS-CoV-2. *Virology Journal*, *17*(1), 138. https://doi.org/10.1186/s12985-020-01395-x

Hu, B., Guo, H., Zhou, P., & Shi, Z.-L. (2021). Characteristics of SARS-CoV-2 and COVID-19. *Nature Reviews Microbiology*, *19*(3), 141–154. https://doi.org/10.1038/s41579-020-00459-7

Inkscape Project. (2020). *Inkscape*. https://inkscape.org

Jared Simpson, & de Borja, R. (2020). *Ncov-tools*. https://github.com/jts/ncov-tools/

Kandeel, M., Ibrahim, A., Fayez, M., & Al-Nazawi, M. (2020). From SARS and MERS CoVs to SARS-CoV-2: Moving toward more biased codon usage in viral structural and nonstructural genes. *Journal of Medical Virology*, *92*(6), 660–666. https://doi.org/10.1002/jmv.25754

Katoh, K., & Standley, D. M. (2013). MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Molecular Biology and Evolution*, *30*(4), 772–780. https://doi.org/10.1093/molbev/mst010

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Meintjes, P., & Drummond, A. (2012). Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics (Oxford, England)*, *28*(12), 1647–1649. https://doi.org/10.1093/bioinformatics/bts199

Kotwa, J. D., Massé, A., Gagnier, M., Aftanas, P., Blais-Savoie, J., Bowman, J., Buchanan, T., Chee, H.-Y., Dibernardo, A., Kruczkiewicz, P., Nirmalarajah, K., Soos, C., Yip, L., Lindsay, L. R., Lung, O., Pickering, B., & Mubareka, S. (2022). *First detection of SARS-CoV-2 infection in Canadian wildlife identified in free-ranging white-tailed deer (Odocoileus virginianus) from southern Québec, Canada* (p. 2022.01.20.476458). bioRxiv. https://doi.org/10.1101/2022.01.20.476458

Kuchipudi, S. V., Surendran-Nair, M., Ruden, R. M., Yon, M., Nissly, R. H., Nelli, R. K., Li, L., Jayarao, B. M., Vandegrift, K. J., Maranas, C. D., Levine, N., Willgert, K., Conlan, A. J. K., Olsen, R. J., Davis, J. J., Musser, J. M., Hudson, P. J., & Kapur, V. (2021). *Multiple spillovers and onward transmission of SARS-CoV-2 in free-living and captive white-tailed deer* (p. 2021.10.31.466677). bioRxiv. https://doi.org/10.1101/2021.10.31.466677

Lam, H. M., Ratmann, O., & Boni, M. F. (2018). Improved Algorithmic Complexity for the 3SEQ Recombination Detection Algorithm. *Molecular Biology and Evolution*, *35*(1), 247–251. https://doi.org/10.1093/molbev/msx263

Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods*, *9*(4), 357–359. https://doi.org/10.1038/nmeth.1923

Legendre, P., & Anderson, M. J. (1999). Distance-Based Redundancy Analysis: Testing Multispecies Responses in Multifactorial Ecological Experiments. *Ecological Monographs*, *69*(1), 1–24. https://doi.org/10.2307/2657192

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., & 1000 Genome Project Data Processing Subgroup. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics (Oxford, England)*, *25*(16), 2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Lu, L., Sikkema, R. S., Velkers, F. C., Nieuwenhuijse, D. F., Fischer, E. A. J., Meijer, P. A., Bouwmeester-Vincken, N., Rietveld, A., Wegdam-Blans, M. C. A., Tolsma, P., Koppelman, M., Smit, L. A. M., Hakze-

van der Honing, R. W., van der Poel, W. H. M., van der Spek, A. N., Spierenburg, M. A. H., Molenaar, R. J., Rond, J. de, Augustijn, M., … Koopmans, M. P. G. (2021). Adaptation, spread and transmission of SARS-CoV-2 in farmed minks and associated humans in the Netherlands. *Nature Communications*, *12*(1), 6802. https://doi.org/10.1038/s41467-021-27096-9

Lung, O., Fisher, M., Nebroski, M., McGregor, G., Schwantje, H., & Joseph, T. (2022). *First whole-genome sequence of Cervid atadenovirus A outside of the United States from an Adenoviral hemorrhagic disease epizootic of black-tailed deer in Canada* (p. 2022.02.10.479879). bioRxiv. https://doi.org/10.1101/2022.02.10.479879

Mannar, D., Saville, J. W., Zhu, X., Srivastava, S. S., Berezuk, A. M., Tuttle, K. S., Marquez, A. C., Sekirov, I., & Subramaniam, S. (2022). SARS-CoV-2 Omicron variant: Antibody evasion and cryo-EM structure of spike protein–ACE2 complex. *Science*. https://doi.org/10.1126/science.abn7760

Memish, Z. A., Mishra, N., Olival, K. J., Fagbo, S. F., Kapoor, V., Epstein, J. H., Alhakeem, R., Durosinloun, A., Al Asmari, M., Islam, A., Kapoor, A., Briese, T., Daszak, P., Al Rabeeah, A. A., & Lipkin, W. I. (2013). Middle East respiratory syndrome coronavirus in bats, Saudi Arabia. *Emerging Infectious Diseases*, *19*(11), 1819–1823. https://doi.org/10.3201/eid1911.131172

Minh, B. Q., Schmidt, H. A., Chernomor, O., Schrempf, D., Woodhams, M. D., von Haeseler, A., & Lanfear, R. (2020). IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Molecular Biology and Evolution*, *37*(5), 1530–1534. https://doi.org/10.1093/molbev/msaa015

Molenaar, R. J., Vreman, S., Hakze-van der Honing, R. W., Zwart, R., de Rond, J., Weesendorp, E., Smit, L. A. M., Koopmans, M., Bouwstra, R., Stegeman, A., & van der Poel, W. H. M. (2020). Clinical and Pathological Findings in SARS-CoV-2 Disease Outbreaks in Farmed Mink (Neovison vison). *Veterinary Pathology*, *57*(5), 653–657. https://doi.org/10.1177/0300985820943535

Mourier, T., Sadykov, M., Carr, M. J., Gonzalez, G., Hall, W. W., & Pain, A. (2021). Host-directed editing of the SARS-CoV-2 genome. *Biochemical and Biophysical Research Communications*, *538*, 35–39. https://doi.org/10.1016/j.bbrc.2020.10.092

Orians, C. M., Schweiger, R., Dukes, J. S., Scott, E. R., & Müller, C. (2019). Combined impacts of prolonged drought and warming on plant size and foliar chemistry. *Annals of Botany*, *124*(1), 41–52. https://doi.org/10.1093/aob/mcz004

O'Toole, Á., Scher, E., Underwood, A., Jackson, B., Hill, V., McCrone, J. T., Colquhoun, R., Ruis, C., Abu-Dahab, K., Taylor, B., Yeats, C., du Plessis, L., Maloney, D., Medd, N., Attwood, S. W., Aanensen, D. M., Holmes, E. C., Pybus, O. G., & Rambaut, A. (2021). Assignment of epidemiological lineages in an emerging pandemic using the pangolin tool. *Virus Evolution*, *7*(2), veab064. https://doi.org/10.1093/ve/veab064

Oude Munnink, B. B., Sikkema, R. S., Nieuwenhuijse, D. F., Molenaar, R. J., Munger, E., Molenkamp, R., van der Spek, A., Tolsma, P., Rietveld, A., Brouwer, M., Bouwmeester-Vincken, N., Harders, F., Hakze-van der Honing, R., Wegdam-Blans, M. C. A., Bouwstra, R. J., GeurtsvanKessel, C., van der Eijk, A. A., Velkers, F. C., Smit, L. A. M., … Koopmans, M. P. G. (2021). Transmission of SARS-CoV-2 on mink farms between humans and mink and back to humans. *Science (New York, N.Y.)*, *371*(6525), 172–177. https://doi.org/10.1126/science.abe5901

Palmer, M. V., Martins, M., Falkenberg, S., Buckley, A., Caserta, L. C., Mitchell, P. K., Cassmann, E. D., Rollins, A., Zylich, N. C., Renshaw, R. W., Guarino, C., Wagner, B., Lager, K., & Diel, D. G. (2021). Susceptibility of White-Tailed Deer (Odocoileus virginianus) to SARS-CoV-2. *Journal of Virology*. https://doi.org/10.1128/JVI.00083-21

Pang, J., & Siu, T. (2022, January 18). Hong Kong to cull 2,000 hamsters after COVID-19 outbreak. *Reuters*. https://www.reuters.com/world/china/hong-kong-orders-hamster-cull-after-covid-19-hits-pets-

2022-01-18/

Patel, H., Varona, S., Monzón, S., Espinosa-Carrasco, J., Heuer, M. L., Underwood, A., Gabernet, G., bot, nf-core, Ewels, P., MiguelJulia, Kelly, S., Erika, Sameith, K., Garcia, M. U., jcurado, & Menden, K. (2022). *nf-core/viralrecon: Nf-core/viralrecon v2.3 - Copper Coatimundi*. Zenodo. https://doi.org/10.5281/zenodo.5974693

Peacock, T. P., Brown, J. C., Zhou, J., Thakur, N., Newman, J., Kugathasan, R., Sukhova, K., Kaforou, M., Bailey, D., & Barclay, W. S. (2022). *The SARS-CoV-2 variant, Omicron, shows rapid replication in human primary nasal epithelial cultures and efficiently uses the endosomal route of entry* (p. 2021.12.31.474653). bioRxiv. https://doi.org/10.1101/2021.12.31.474653

Pedersen, B. S., & Quinlan, A. R. (2018). Mosdepth: Quick coverage calculation for genomes and exomes. *Bioinformatics (Oxford, England)*, *34*(5), 867–868. https://doi.org/10.1093/bioinformatics/btx699

Public Health Ontario. (2022). *SARS-CoV-2 Whole Genome Sequencing in Ontario,* (Weekly Epidemiological Summary). Public Health Ontario.

Rambaut, A., Holmes, E. C., O'Toole, Á., Hill, V., McCrone, J. T., Ruis, C., du Plessis, L., & Pybus, O. G. (2020). A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. *Nature Microbiology*, *5*(11), 1403–1407. https://doi.org/10.1038/s41564-020-0770-5

Ringlander, J., Fingal, J., Kann, H., Prakash, K., Rydell, G., Andersson, M., Martner, A., Lindh, M., Horal, P., Hellstrand, K., & Kann, M. (2022). Impact of ADAR-induced editing of minor viral RNA populations on replication and transmission of SARS-CoV-2. *Proceedings of the National Academy of Sciences*, *119*(6). https://doi.org/10.1073/pnas.2112663119

Shan, K.-J., Wei, C., Wang, Y., Huan, Q., & Qian, W. (2021). Host-specific asymmetric accumulation of mutation types reveals that the origin of SARS-CoV-2 is consistent with a natural process. *Innovation (New York, N.Y.)*, *2*(4), 100159. https://doi.org/10.1016/j.xinn.2021.100159

Shriner, S. A., Ellis, J. W., Root, J. J., Roug, A., Stopak, S. R., Wiscomb, G. W., Zierenberg, J. R., Ip, H. S., Torchetti, M. K., & DeLiberto, T. J. (2021). SARS-CoV-2 Exposure in Escaped Mink, Utah, USA. *Emerging Infectious Diseases Journal*, *27*(3). https://doi.org/10.3201/eid2703.204444

Shu, Y., & McCauley, J. (2017). GISAID: Global initiative on sharing all influenza data – from vision to reality. *Eurosurveillance*, *22*(13), 30494. https://doi.org/10.2807/1560-7917.ES.2017.22.13.30494

Shuai, H., Chan, J. F.-W., Yuen, T. T.-T., Yoon, C., Hu, J.-C., Wen, L., Hu, B., Yang, D., Wang, Y., Hou, Y., Huang, X., Chai, Y., Chan, C. C.-S., Poon, V. K.-M., Lu, L., Zhang, R.-Q., Chan, W.-M., Ip, J. D., Chu, A. W.-H., … Chu, H. (2021). Emerging SARS-CoV-2 variants expand species tropism to murines. *EBioMedicine*, *73*. https://doi.org/10.1016/j.ebiom.2021.103643

Simmonds, P., & Ansari, M. A. (2021). Extensive C->U transition biases in the genomes of a wide range of mammalian RNA viruses; potential associations with transcriptional mutations, damage- or host-mediated editing of viral RNA. *PLOS Pathogens*, *17*(6), e1009596. https://doi.org/10.1371/journal.ppat.1009596

Starr, T. N., Greaney, A. J., Hilton, S. K., Ellis, D., Crawford, K. H. D., Dingens, A. S., Navarro, M. J., Bowen, J. E., Tortorici, M. A., Walls, A. C., King, N. P., Veesler, D., & Bloom, J. D. (2020). Deep Mutational Scanning of SARS-CoV-2 Receptor Binding Domain Reveals Constraints on Folding and ACE2 Binding. *Cell*, *182*(5), 1295-1310.e20. https://doi.org/10.1016/j.cell.2020.08.012

Tan, C. C. S., Lam, S. D., Richard, D., Owen, C., Berchtold, D., Orengo, C., Nair, M. S., Kuchipudi, S. V., Kapur, V., Dorp, L. van, & Balloux, F. (2022). Transmission of SARS-CoV-2 from humans to animals and potential host adaptation. *BioRxiv*, 2020.11.16.384743. https://doi.org/10.1101/2020.11.16.384743

Tsueng, G., Mullen, J., Alkuzweny, M., Cano, M. A., Rush, B., Haag, E., Curators, O., Latif, A. A., Zhou, X., Qian, Z., Andersen, K. G., Wu, C., Su, A. I., Gangavarapu, K., & Hughes, L. D. (2022). *Outbreak.info Research*

Library: A standardized, searchable platform to discover and explore COVID-19 resources and data (p. 2022.01.20.477133). bioRxiv. https://doi.org/10.1101/2022.01.20.477133

Turakhia, Y., Thornlow, B., Hinrichs, A. S., De Maio, N., Gozashti, L., Lanfear, R., Haussler, D., & Corbett-Detig, R. (2021). Ultrafast Sample placement on Existing tRees (UShER) enables real-time phylogenetics for the SARS-CoV-2 pandemic. *Nature Genetics*, *53*(6), 809–816. https://doi.org/10.1038/s41588-021-00862-7

Varabyou, A., Pockrandt, C., Salzberg, S. L., & Pertea, M. (2021). Rapid detection of inter-clade recombination in SARS-CoV-2 with Bolotie. *Genetics*, *218*(3), iyab074. https://doi.org/10.1093/genetics/iyab074

Wei, C., Shan, K.-J., Wang, W., Zhang, S., Huan, Q., & Qian, W. (2021). Evidence for a mouse origin of the SARS-CoV-2 Omicron variant. *Journal of Genetics and Genomics*. https://doi.org/10.1016/j.jgg.2021.12.003

Wright, F. (1990). The 'effective number of codons' used in a gene. *Gene*, *87*(1), 23–29. https://doi.org/10.1016/0378-1119(90)90491-9

Wu, F., Zhao, S., Yu, B., Chen, Y.-M., Wang, W., Song, Z.-G., Hu, Y., Tao, Z.-W., Tian, J.-H., Pei, Y.-Y., Yuan, M.-L., Zhang, Y.-L., Dai, F.-H., Liu, Y., Wang, Q.-M., Zheng, J.-J., Xu, L., Holmes, E. C., & Zhang, Y.-Z. (2020). A new coronavirus associated with human respiratory disease in China. *Nature*, *579*(7798), 265–269. https://doi.org/10.1038/s41586-020-2008-3

Yen, H.-L., Sit, T. H., Brackman, C. J., Chuk, S. S., Cheng, S. M. S., Gu, H., Chang, L. D., Krishnan, P., Ng, D. Y., Liu, G. Y., Hui, M. M., Ho, S. Y., Tam, K. W., Law, P. Y., Su, W., Sia, S. F., Choy, K.-T., Cheuk, S. S., Lau, S. P., … Poon, L. L. (2022). *Transmission of SARS-CoV-2 (Variant Delta) from Pet Hamsters to Humans and Onward Human Propagation of the Adapted Strain: A Case Study* (SSRN Scholarly Paper ID 4017393). Social Science Research Network. https://doi.org/10.2139/ssrn.4017393

Yu, G., Smith, D. K., Zhu, H., Guan, Y., & Lam, T. T.-Y. (2017). ggtree: An r package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods in Ecology and Evolution*, *8*(1), 28–36. https://doi.org/10.1111/2041-210X.12628

Zhan, S. H., Alamouti, S. M., Daneshpajouh, H., Kwok, B. S., Lee, M.-H., Khattra, J., Houck, H. J., & Rand, K. H. (2021). Target capture sequencing of SARS-CoV-2 genomes using the ONETest Coronaviruses Plus. *Diagnostic Microbiology and Infectious Disease*, *101*(3), 115508. https://doi.org/10.1016/j.diagmicrobio.2021.115508

Zhou, J., Peacock, T. P., Brown, J. C., Goldhill, D. H., Elrefaey, A. M. E., Penrice-Randal, R., Cowton, V. M., Lorenzo, G. D., Furnon, W., Harvey, W. T., Kugathasan, R., Frise, R., Baillon, L., Lassaunière, R., Thakur, N., Gallo, G., Goldswain, H., Donovan-Banfield, I., Dong, X., … Barclay, W. S. (2022). Mutations that adapt SARS-CoV-2 to mink or ferret do not increase fitness in the human airway. *Cell Reports*, *38*(6). https://doi.org/10.1016/j.celrep.2022.110344

Supplementary Material

Figures

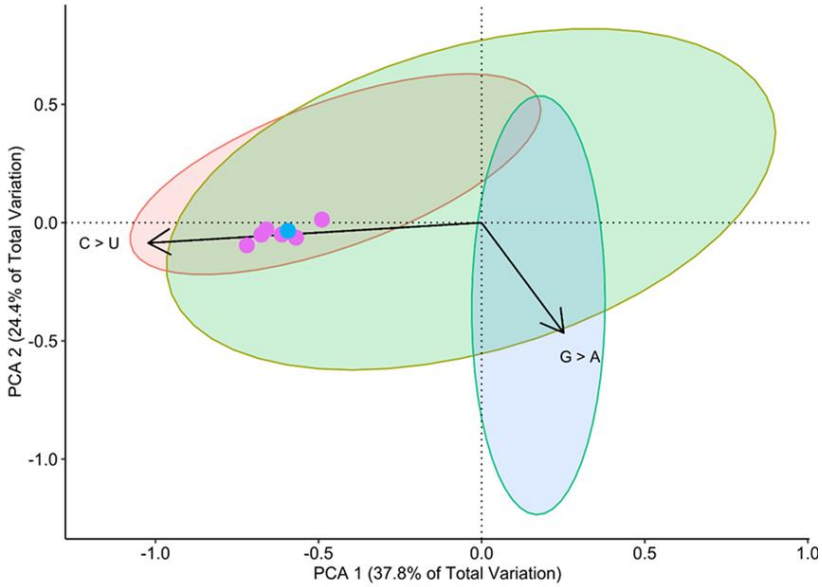

Figure S1. Principal Components Analysis of the mutational spectra of SARS-CoV-2 genomes isolated from different hosts within Clade 20C. The first two components account for 62.2% of the variation between the samples. Variation in the spectra along the first principal component associated with changes in the frequency of C>U. Samples appear to be spread along the first component by host-type, although this effect is not as strong as in Figure 6. Ontario WTD samples (pink) and a human sample from Ontario (blue) appear close together in the projection, suggesting that they share a very similar mutation spectrum.

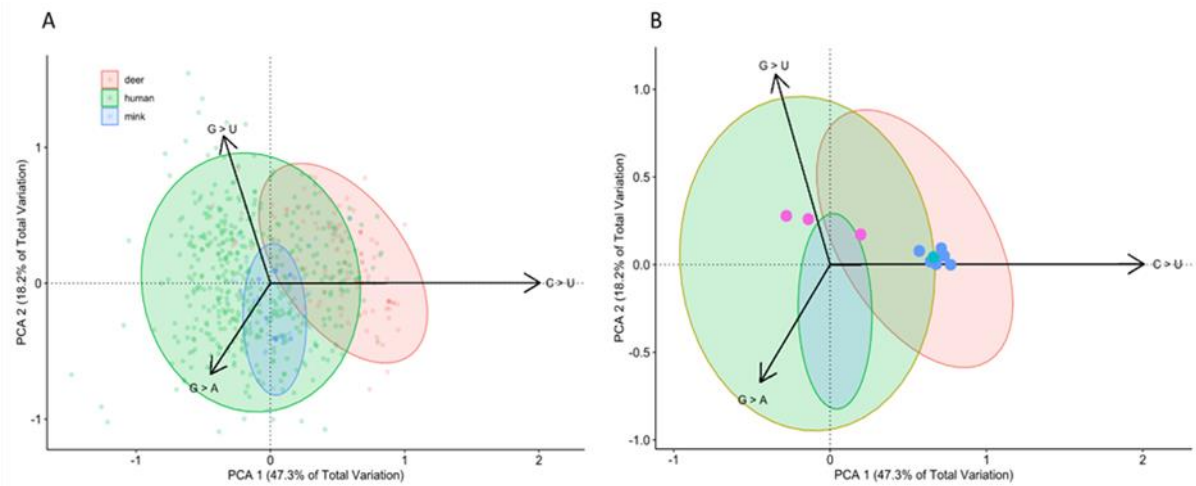

Figure S2. Principal Components Analysis of the mutational spectra from different SARS-CoV-2 variants. The first two components account for 65.5% of the variation between the samples. Variation in the spectra along the first principal component associated with changes in the frequency of C>U mutations. Interestingly, samples along this component are also differentiated by host-type. Variation along the second principal component reflect changes in the frequency of G>U and G>A mutations. (A) The PCA biplot showing a projection of all samples. (B) A simplified version of plot A highlighting the positions of WTD samples isolated from Ontario (Blue) and Quebec (Pink). A human isolate from Ontario appears in

the same area of the plot (teal) as other Ontario WTD samples. Arrows are scaled to 30% of their original size to create a cleaner plot.

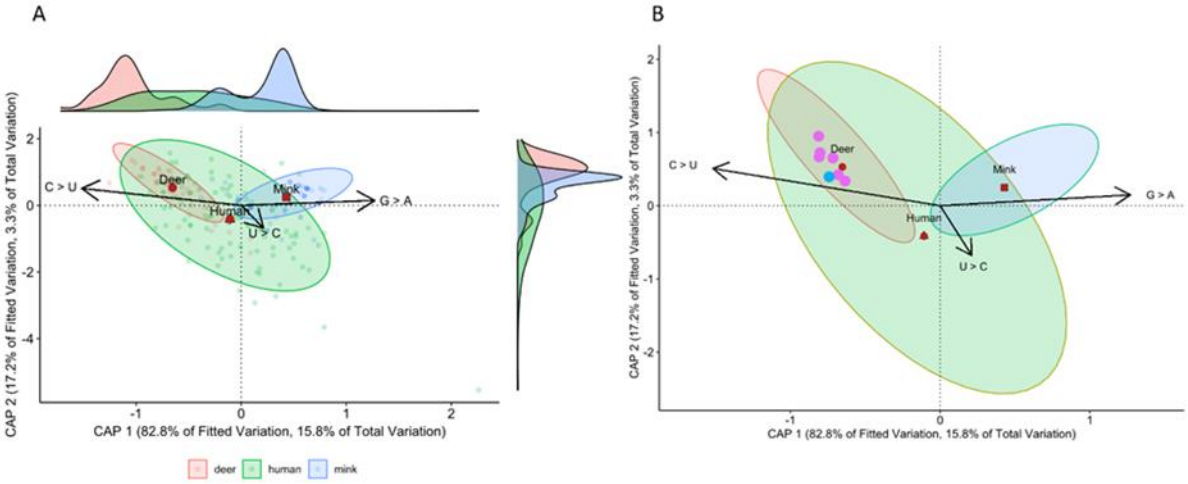

Figure S3. A redundancy analysis illustrating the relative positions of all samples from each host in the constrained space. Each set of samples collected from human, mink, and deer hosts is surrounded by the 95% confidence ellipse. Host-type explains 19.1% of the variation in the mutational spectra. The distribution of each set of samples along each axis within the constrained space is shown in (A). The location of Ontario WTD samples (pink) and the human isolate (teal) within the constrained space is shown in (B). The mutational spectra of these sequences is consistent with that from other white-tailed deer isolates.

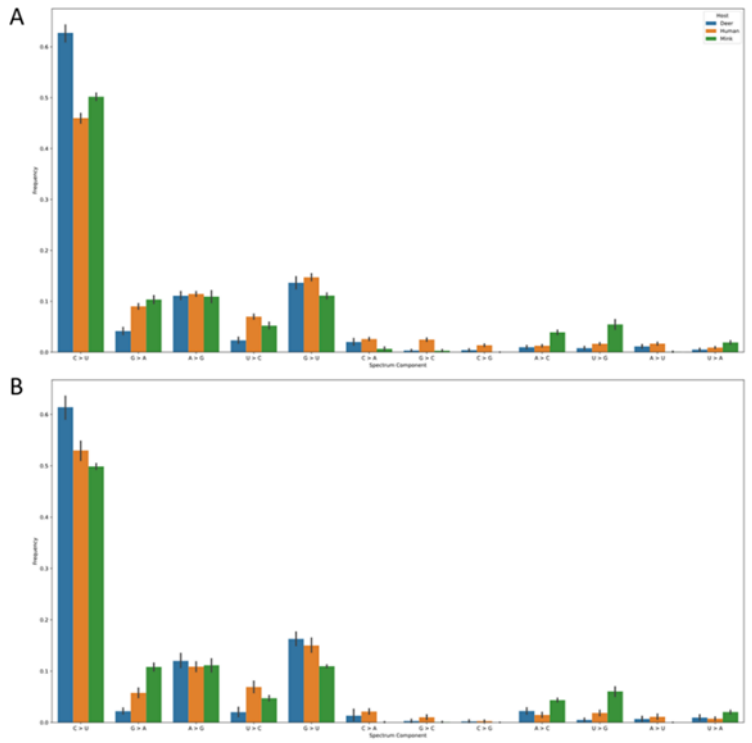

Figure S4. The frequency of different components of the mutational spectra from all samples (A) and only those found in Clade 20C (B).

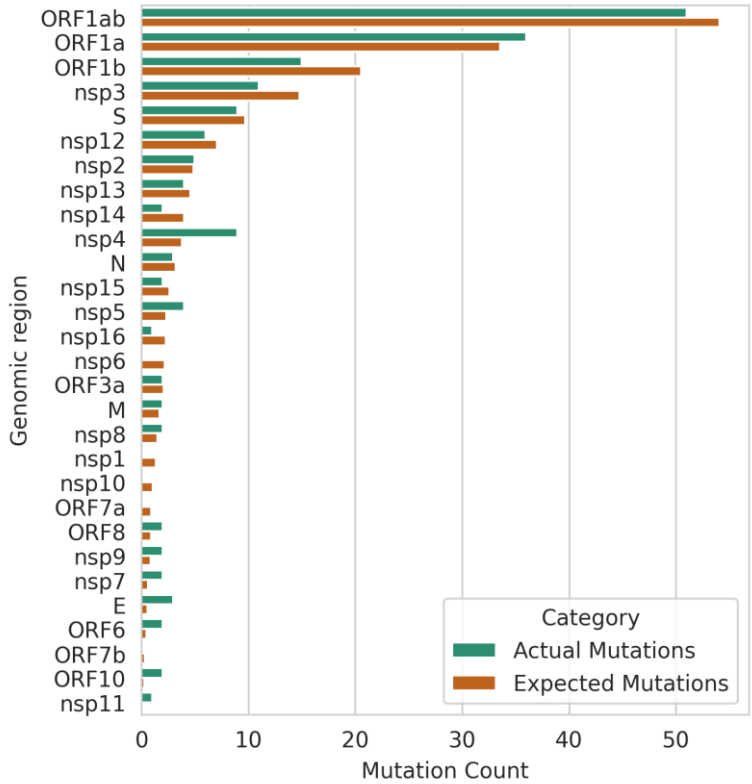

Figure S5: Observed vs expected counts within shared ON WTD clade mutations for each gene/product.

Figure S6: Usher based phylogenetic tree with Ontario WTD high quality (n=5) and partial (n=2) genomes and the 2022-02-21 USCS Usher build (see .nwk file).

Tables

Table S1: Sampling data breaking down % positivity of PCR/Serology/Genome and relationship between samples. Sample IDs used in the manuscript are the last 4 digits of the WILD-CoV Sample ID. * Samples not selected for sequencing with combined capture enrichment and ARTICv4 approach based on initial quality of ARTICv4 sequencing. [+] Human-derived sample sequenced by PHOL as part of public health surveillance. RPLN: Retrophyarngeal Lymph Nodes, RT-PCRs at SRI (UTR and E gene Ct<40) and CFIA (E and N2 gene Ct<36) are indicated in that order. Positivity is indicated by +, inconclusive results with +?, and negative results with -.

Table S2: Summary of mutations within Ontario WTD clade (with associated human sequence) and their distribution across GISAID sequences, VOC, animal-derived viral sequences, and related Michigan mink sequences.

Table S3: Poor coverage regions in Ontario WTD clade (with associated human sequence) sequenced with combined ARTIC V4 and capture enrichment (ONETest) data.

Table S4: Results of a distance-based Welch MANOVA investigating differences in the mutation spectrum between hosts within Nextstrain Clade 20C.

Table S5. Summary of codon usage bias analysis results across SARS-CoV-2 from WTD (including the Ontario lineage) and other cervid viruses.

Table S6. Acknowledgement table for sequences used from GISAID for phylogenetic and mutational signature analyses.