Peer Review File

Manuscript Title: SARS-CoV-2 infection in free-ranging white-tailed deer

Reviewer Comments & Author Rebuttals

Reviewer Reports on the Initial Version:

Referee #1

The manuscript "SARS-CoV-2 infection in free-ranging white-tailed deer (Odocoileus virginianus)" provides a clear and thorough description of detection and genetic characterization of SARS-COV-2 in this novel host with an accompanying discussion of the importance of this reverse transmission and the rare mutations that are being fixed in this new host. Of course, the importance is the possibility for adaptation in new host species and the potential for genetic drift in those populations to generate viruses that could then reinfect humans – effectively replicating what happened with the 2009 influenza pandemic. We cannot overstate the importance of these findings, or the risk posed by this and other human to animal transmissions. This is an simple yet outstanding and extremely timely study that reminds us of the multiple challenges that are posed by not stopping an emerging infection. It should act as a rallying call for a One Health approach not just for the prediction of new pandemics but in the active management and pandemics and the many associated reverse zoonotic infections.

Technically the paper is very sound, I have a few simple questions/clarifications below. The conclusions are clear, well presented and supported by the data. The writing is lovely and beautifully clear.

Minor comments.
1. It would be helpful to at least know how many primer pairs were used in the PCR protocol so we know amplicon length.
2. What viruses were sequenced with Illumina vs ONT?
3. How were the animals transported to the sample site after culling? Is there any chance of cross contamination, particularly for those samples that formed clusters?

Referee #2

This manuscript reports detection of SARS-CoV-2 virus RNA by RT-PCR in white tailed deer in 9 of 9 different sampling sites within the State of Ohio, during the winter of 2020/21. Positive samples were collected on 17 of 18 sampling dates. The overall prevalence of positive RT-PCR results was 35.8% with prevalence ranging from 13.5-70% for different sampling sites. Fourteen full genome virus sequences were obtained and the analysis of the sequence data indicates at least 6 separate spill-over events from humans to deer and also provides convincing evidence of transmission within deer, with unusual mutations appearing within the deer population.

Specific comments and queries:
The very high detection rate of SARS-CoV-2 RNA in the deer populations is remarkable. Do the authors have any data on the sero-prevalence in deer at this time.

Although the reviewer does not doubt the validity of the RT-PCR detection data, it may be useful to spell out what measures were in place to prevent cross contamination, both at sampling sites and in the laboratory. For example, were positive samples re-extracted and retested for confirmation?

Line 130 onwards: The authors comment of a number of mutations found in deer in clade B.1.596

viruses at site 1. Can the authors clarify of ALL these mutations were found in all viruses? Or was there a progressive accumulation of mutations? From the way it is written, it appears to be the former scenario but it would be important to be absolutely clear on this point.

Line 138 onwards: Amino acid substitution of position 484 has been reported in a number of variants of concern or interest (Beta, Gamma, Eta, Iota, Mu). It is possible that this E484D mutation appeared in a human as a result of partial immune escape which transmitted to deer or it may have appeared within deer as a result of immune pressure accumulating within the deer population. Can the authors make these two possibilities clear. Can the authors clarify if this particular E484D mutation has ever been seen in humans.

Line 172-3. Viable virus has been detected in wastewater? Viable virus can only be confirmed by virus isolation in culture. A quick perusal of one of the two references cited did not mention virus culture. Can the authors clarify this. In this reviewers experience, although virus RNA can be readily detected in human stool, culture of virus is rare, though reported occasionally. I am not sure if the implication that faecal virus was the likely source of this spillover to deer is really significant. Even the other indirect possibilities for transmission such as trash etc is probably not as likely as direct interaction between humans and deer. Does direct human-deer contact occur? Such as via the general public feeding and petting deer etc?

The initial detection of serological evidence of SARS-CoV-2 infection in white tailed deer and this report of detecting and sequencing virus RNA from deer are likely facilitated by the fact that these deer are being culled for population control, thus facilitating sample collection for investigation. It is therefore an open question of how many other species which are less accessible to sampling may be actually getting infected via spill over infection. Thus, while tailed deer may only be the tip of an iceberg. This point is worth making explicit.

Referee #3

This manuscript describes results of a survey of SARS COV 2 among deer in 9 locations in a region in Ohio. The animals were captured as part of a deer management program, which involved baiting. The authors found a high prevalence of PCR positive animals, with prevalences varying per site but up to a remarkable 70%. A small proportion of the positives were sequenced (14 out of 129 positives). The analysis of the genomes in comparison with viruses from humans found evidence for repeated introductions of human viruses into the deer population, and for deer-to-deer transmission in some of the sites. Phylogenetic analysis and the identification of rare mutations in the deer genomes compared to the human genetic sequences suggested that the virus had been circulating for some months in some locations.

The results add to a worrisome trend of spillbacks of SARS COV 2 into wild animals and the authors do discuss the concern that this may reflect establishment of SARS COV 2 in an animal reservoir.

Major comments:

The manuscript would benefit from a more detailed description of the deer population and the interactions specific to the region. Is there any deer farming, are deer being fed, what are interactions between deer and humans in the urban areas? Is anything known about potential other interactions between humans and deer, for instance through (infected) cats or dogs? Similarly, a more detailed description and discussion of the observations in that context is needed. A point prevalence of 70% is difficult to reconcile with what is currently known about SARS COV 2 infection in most animal species. Here, infection is self-limiting, so finding such a high prevalence would suggest either some type of bias in the animals caught, or recent outbreaks. A question that comes to mind is whether the baiting may contribute to introductions.

The relatively small proportion of genomes sequenced (11%) is a weakness of the study, and I would strongly recommend expanding the number of genomes. Getting full genomes may be challenging but assuming that Ct 30 is a cut-off for successful whole genome sequencing in most labs, there is room for improvement.

Minor comments

Page 3, line 49: free ranging mink infected have also been observed elsewhere (US, Netherlands, possibly more). Needs updating.

Line 323: describes that deer were baited. How was that done, and is the health status of persons involved in the process known? Could this have been a source of introduction? Is feeding part of deer management?

Line 353: sequencing protocol is available on request. This should be included in the documentation.

Line 424: Estimation of proportion of sequences per week assumes that the sequencing selection is random. Is that indeed the case? How is the selection of samples for sequencing for the human COVID19 genomic surveillance organised? Needs a description.

Results section from line 91 :
14 genomes were sequenced. It is unclear how they were distributed over the sites, and the number of samples that would allow sequencing based on Ct values – assuming a cut off of Ct 30 in the E gene PCR, is close to 50. Why were not more samples sequenced? Surely this would have given a much deeper picture. For instance, one could hypothesis that the process of the study itself may have triggered an outbreak on sites 1 and 6 given the baiting involved in getting access to the animals, and the increased prevalence at the second sampling time point.

Line 104: this is where the relatively sparse coverage of sequencing really is too bad, because the question is whether this is simply explained by undersampling and possibly more recent introduction of the other lineages.

Author Rebuttals to Initial Comments:

December 9, 2021

We would like to thank the referees for their thoughtful reviews and the suggestions to improve our manuscript. We have addressed each one of the reviewers' comments individually in blue text.

1.Please reduce the title to 75 characters or less, including spaces. (I suggest just removing the latin name of the deer.)

We have changed the title as suggested.

2. Please include references in the abstract.

We have included five new references in the abstract.

3. Is ‘figure 3’ is mislabeled as ‘figure 4”? There is no figure 3 included in this version of the ms, but it is referenced in the text, and there are no references to a fig 4.

Yes. It has now been relabeled as Figure 3

4. Please remove figures from the main article (leaving figure titles & legends) and resupply as separate, individual .eps, .ai, .ps, .pdf, or .ppt files

The figures have been removed from the article and have been uploaded as separate .pdf files.

5. Extended data tables should also be removed from the main article (leaving the footnotes) and resupplied as separate, individual .eps, .jpeg, or .tif files.

The extended data tables have been removed from the article and have been uploaded as separate .tif files.

6 Please clarify the source of the human and stag silhouettes in some of the figures and whether permissions are required for publication.

We have replaced the human and stag silhouettes with versions that do not require attribution.

7. Some of the text in figure 1 (the word ‘kilometers’ as well as the map attribution) is slightly too small at double-column width. Please ensure that text is the equivalent size to a readable 5pt Arial font when the figures are at the final size.

The text in Figure 1 has been resized to improve readability.

8. Figures 2 and 3 are too tall at the necessary double-column width. Please ensure that these figures do not exceed 17cm in height.

Figures 2 and 3 have been resized to meet the guidelines.

Referee #1 (Remarks to the Author):
The manuscript “SARS-CoV-2 infection in free-ranging white-tailed deer (Odocoileus virginianus)” provides a clear and thorough description of detection and genetic characterization of SARS-COV-2 in this novel host with an accompanying discussion of the importance of this reverse transmission and the rare mutations that are being fixed in this new host. Of course, the importance is the possibility for adaptation in new host species and the potential for genetic drift in those populations to generate viruses that could then reinfect humans – effectively replicating what happened with the 2009 influenza pandemic. We cannot overstate the importance of these findings, or the risk

posed by this and other human to animal transmissions. This is an simple yet outstanding and extremely timely study that reminds us of the multiple challenges that are posed by not stopping an emerging infection. It should act as a rallying call for a One Health approach not just for the prediction of new pandemics but in the active management and pandemics and the many associated reverse zoonotic infections.

Technically the paper is very sound, I have a few simple questions/clarifications below. The conclusions are clear, well presented and supported by the data. The writing is lovely and beautifully clear.

Minor comments.
1. It would be helpful to at least know how many primer pairs were used in the PCR protocol so we know amplicon length.

The ARTIC v3 primer panel uses 109 total primer pairs in two separate pools, to avoid overlap. The amplicons are approximately 400 bps. As suggested by Reviewer 3, we have now included a reference for the sequencing protocol used by NVSL. The NVSL protocol uses six primer pools averaging 550 bps with 100 bp overlaps.

2. What viruses were sequenced with Illumina vs ONT?

All viruses, regardless of submitting laboratory, were sequenced using Illumina. We scanned the manuscript for any erroneous references to Oxford Nanopore sequencing, but found none.

3. How were the animals transported to the sample site after culling? Is there any chance of cross contamination, particularly for those samples that formed clusters?

We agree that this is critical information and have added it to the Methods section:

"Harvest occurred at locations that were baited with whole kernel corn for up to two weeks prior to each culling session, and additional deer were harvested opportunistically when they were observed away from the bait on a culling session day. In the field, once a deer was harvested, the head was wrapped in a plastic bag and an identification tag was attached to a leg. Each day of the program, harvested deer carcasses were transported to a central processing point where samples were collected. Sample collectors wore gloves and a facemask. A nasal swab was collected from each deer and placed into a tube with brain heart infusion broth (BHIB). After collection, samples were immediately chilled on ice packs then transferred into a – 80°C freezer within 12 h where they remained until testing was initiated. Samples were collected post-mortem, which was exempt from The Ohio State University Institutional Animal Care and Use Committee."

Referee #2 (Remarks to the Author):

This manuscript reports detection of SARS-CoV-2 virus RNA by RT-PCR in white tailed deer in 9 of 9

different sampling sites within the State of Ohio, during the winter of 2020/21. Positive samples were collected on 17 of 18 sampling dates. The overall prevalence of positive RT-PCR results was 35.8% with prevalence ranging from 13.5-70% for different sampling sites. Fourteen full genome virus sequences were obtained and the analysis of the sequence data indicates at least 6 separate spill-over events from humans to deer and also provides convincing evidence of transmission within deer, with unusual mutations appearing within the deer population.

Specific comments and queries:
The very high detection rate of SARS-CoV-2 RNA in the deer populations is remarkable. Do the authors have any data on the sero-prevalence in deer at this time.

Unfortunately, we do not have seroprevalence data at this time. The only seroprevalence data for free-ranging white-tailed deer that we are currently aware of are from Chandler et al. https://doi.org/10.1073/pnas.2114828118. While that paper does not have data from deer in Ohio, the study found seroprevalence of 40% in free-ranging white-tailed deer across several states, including two that border Ohio (Pennsylvania and Michigan). We are prospectively collecting sera from white-tailed deer in Ohio going forward.

Although the reviewer does not doubt the validity of the RT-PCR detection data, it may be useful to spell out what measures were in place to prevent cross contamination, both at sampling sites and in the laboratory. For example, were positive samples re-extracted and retested for confirmation?

The safeguards taken during sampling were described in our response above to Reviewer 1.

With regards to reducing the chance of cross-contamination of samples during laboratory testing, appropriate standard techniques were applied during transfer and working with samples and extracted nucleic acid to avoid cross contamination. A subset of original samples were selected for re-extraction and retesting at Ohio State University to verify the original results. NVSL (USDA) performed extraction and subsequent testing on original material provided by OSU, not on their extracted material. 75 of the 76 samples tested positive at NVSL. When working with SARS-CoV-2 samples, NVSL personnel wear powered air purifying respirators (PAPRs) and work with the samples inside class 2A biosafety cabinets. Only one sample is open in the hood at a time.

Our confidence in excluding laboratory contamination is increased due to the fact that the samples were not processed in the OSU laboratory until months after they were collected in deer. The time of sample testing, different SARS-CoV-2 variants were circulating in humans, making it easier to flag more recent contamination events.

Line 130 onwards: The authors comment of a number of mutations found in deer in clade B.1.596 viruses at site 1. Can the authors clarify of ALL these mutations were found in all viruses? Or was there a progressive accumulation of mutations? From the way it is written, it appears to be the former scenario but it would be important to be absolutely clear on this point.

All six clade-defining mutations were observed in all seven Site 1 sequences. We have added text to clarify this. Additionally, we comment on one spike mutation in B.1.596 viruses from single deer at site 1 (E484D). We have changed the text from "A B.1.596 virus from site 1", to "A single B.1.596 virus from site 1" to make it abundantly clear that we are only talking about one virus.

Line 138 onwards: Amino acid substitution of position 484 has been reported in a number of variants of concern or interest (Beta, Gamma, Eta, Iota, Mu). It is possible that this E484D mutation appeared in a human as a result of partial immune escape which transmitted to deer or it may have appeared within deer as a result of immune pressure accumulating within the deer population. Can the authors make these two possibilities clear. Can the authors clarify if this particular E484D mutation has ever been seen in humans.

Yes, this mutation has been seen in humans, albeit at rare frequency. E484D has been seen in 201 sequences from humans; 71 were from US. While E484D is not common, E484K is highly prevalent in human samples. We have added new text in the Results section:

"The E484D substitution has only been detected in 201 SARS-CoV-2 sequences from humans globally, 71 of which were in the United States, but none of the B.1.596 viruses in humans that were most closely related to the deer virus have this mutation. It is therefore impossible to differentiate if the E484D mutation arose in an unsampled human virus and was transmitted to deer or arose de novo in deer."

Line 172-3. Viable virus has been detected in wastewater? Viable virus can only be confirmed by virus isolation in culture. A quick perusal of one of the two references cited did not mention virus culture. Can the authors clarify this. In this reviewers experience, although virus RNA can be readily detected in human stool, culture of virus is rare, though reported occasionally. I am not sure if the implication that faecal virus was the likely source of this spillover to deer is really significant. Even the other indirect possibilities for transmission such as trash etc is probably not as likely as direct interaction between humans and deer. Does direct human-deer contact occur? Such as via the general public feeding and petting deer etc?

Thank you for pointing this out. Viable virus has been detected in stool and experimentally in wastewater (Bivins et al 2020) but not under natural conditions. (The Bivins article spiked virus into wastewater and infectivity (culture in Vero E6 cells) was tested over time and under differing environmental conditions.) We have now clarified this in the text: "SARS-CoV-2 RNA has also been detected in wastewater[28,29] and urban runoff[30]; although, infectivity of SARS-CoV-2 from these sources is unknown."

Deer-human contact does occur during public feeding or contact in a wildlife hospital, but is rare. Given the number of human-to-deer spillovers based on sequencing data, and the season (winter) an indirect route seems more feasible; although we cannot rule out other possibilities.

The initial detection of serological evidence of SARS-CoV-2 infection in white tailed deer and this report of detecting and sequencing virus RNA from deer are likely facilitated by the fact that these deer are being culled for population control, thus facilitating sample collection for investigation. It is

therefore an open question of how many other species which are less accessible to sampling may be actually getting infected via spillover infection. Thus, white-tailed deer may only be the tip of an iceberg. This point is worth making explicit.

We concur and have added several final lines at the end of the Discussion to make this point explicit.

"Moreover, it is worth noting that white-tailed deer are a relatively convenient surveillance target because of their abundance and accessibility. The detection of SARS-CoV-2 in free-ranging white-tailed deer naturally raises the question whether less accessible species are also being infected through viral spillover from humans and calls for broader surveillance efforts."

Referee #3 (Remarks to the Author):

This manuscript describes results of a survey of SARS COV 2 among deer in 9 locations in a region in Ohio. The animals were captured as part of a deer management program, which involved baiting. The authors found a high prevalence of PCR positive animals, with prevalences varying per site but up to a remarkable 70%. A small proportion of the positives were sequenced (14 out of 129 positives). The analysis of the genomes in comparison with viruses from humans found evidence for repeated introductions of human viruses into the deer population, and for deer-to-deer transmission in some of the sites. Phylogenetic analysis and the identification of rare mutations in the deer genomes compared to the human genetic sequences suggested that the virus had been circulating for some months in some locations.

The results add to a worrisome trend of spillbacks of SARS COV 2 into wild animals and the authors do discuss the concern that this may reflect establishment of SARS COV 2 in an animal reservoir.

Major comments:

The manuscript would benefit from a more detailed description of the deer population and the interactions specific to the region. Is there any deer farming, are deer being fed, what are interactions between deer and humans in the urban areas? Is anything known about potential other interactions between humans and deer, for instance through (infected) cats or dogs? Similarly, a more detailed description and discussion of the observations in that context is needed. A point prevalence of 70% is difficult to reconcile with what is currently known about SARS COV 2 infection in most animal species. Here, infection is self-limiting, so finding such a high prevalence would suggest either some type of bias in the animals caught, or recent outbreaks. A question that comes to mind is whether the baiting may contribute to introductions.

We have incorporated the available information on the deer population that was studied in the Discussion section:

"Estimates of deer density in and around our sites range from approximately 8/km2 to upwards of 30/km2. There are no deer farms in the study area and public feeding of deer is prohibited. There is

ample forage available around urban and suburban residences in gardens and plantings, drawing deer into close proximity with humans and their companion animals."

Additionally, we added text in the Discussion section to acknowledge the potential of baiting to facilitate SARS-CoV-2 spread in these deer populations:

"It is unclear if baiting the deer prior to harvest contributed the increased frequency of SARS-CoV-2 in this study, but concentrating deer with bait could have potentially facilitated pathogen transmission through a population. However, baiting is regularly used in deer management programs and the practice is commonly employed by deer hunters, which makes understanding the effect of baiting on SARS-CoV-2 transmission in free-ranging deer paramount for future studies."

The relatively small proportion of genomes sequenced (11%) is a weakness of the study, and I would strongly recommend expanding the number of genomes. Getting full genomes may be challenging but assuming that Ct 30 is a cut-off for successful whole genome sequencing in most labs, there is room for improvement.

We concur that 11% is a low sequencing rate, but it reflects sample degradation that is common in convenience sampling in field settings, not lack of effort in the lab. NVSL attempted repeated WGS on the PCR positive samples submitted to NVSL, this included attempts which also integrated a host depletion step. Resequencing with the original method and resequencing with the MagMax Core kit were attempted, however host nucleic acid contamination was significant. Fourteen samples were subjected to additional host depletion, which significantly reduced the proportion of host nucleic acid but did not improve genome coverage of the sequences. The genomes in the study represent the sequences which were of sufficient quality for additional analysis and characterization. OSU also attempted WGS on a subset of the samples. Going forward, samples will be collected from white-tailed deer specifically for SARS-CoV-2 surveillance and will be processed more quickly.

Minor comments

Page 3, line 49: free ranging mink infected have also been observed elsewhere (US, Netherlands, possibly more). Needs updating.

These lines now read: "The largest outbreaks have been observed on mink farms in Denmark and the Netherlands. Detection of SARS-CoV-2 by PCR in free-ranging wildlife has been limited to small numbers of mink in Spain and Utah (USA), which purportedly escaped from a nearby farm[4 5]." While there is anecdotal evidence for SARS-CoV-2 detections in free-ranging mink in the Netherlands based on serology, we could not find a published study and therefore excluded it.

Line 323: describes that deer were baited. How was that done, and is the health status of persons

involved in the process known? Could this have been a source of introduction? Is feeding part of deer management?

Baiting is commonly part of deer management programs. Bait sites were baited with whole kernel corn that was stored in bags until use. Sites were divided into regions and a single staff member baited sites at each region (Region 1 = sites 1, 6, 8; Region 2 = sites 3, 4; Region 3 = sites 2, 5, 7; Region 4= Site 9). We do not have information on the health of the staff members who were baiting; however, the distinct viral introductions at each site indicates there was not a common source that bridged sites.

Line 353: sequencing protocol is available on request. This should be included in the documentation.

We have added a reference for this protocol Emerg Infect Dis. 2020 Oct; 26(10): 2401–2405. doi: 10.3201/eid2610.201800.

Line 424: Estimation of proportion of sequences per week assumes that the sequencing selection is random. Is that indeed the case? How is the selection of samples for sequencing for the human COVID19 genomic surveillance organised? Needs a description.

We are aware of sampling biases in US genomic surveillance data. For this reason our initial analysis used only Ohio sequences that were labeled as "baseline surveillance" in GISAID, which excludes sequences collected from targeted groups or settings for specific research questions. Approximately one-third of Ohio's sequences were labeled as "baseline" from this time period. However, baseline surveillance did not begin in earnest until 2021, so to maximize our dataset we compared lineage proportions observed in the small baseline dataset against lineage proportions observed in the larger complete dataset. Because the proportions were very similar, we included the larger dataset in the final figure. We explain this in the Methods section as follows:

"To further minimize biases only sequences categorized in the GISAID submission as obtained using a "baseline surveillance" sampling strategy were included in the original analysis. The dataset was further trimmed to include only submissions with complete collection dates and sufficient coverage to assign a Pango lineage, resulting in a final dataset of 9,947 sequences from Ohio. For simplicity sub-lineages of B.1.617.2 (e.g., AY.3) were consolidated into the Delta category and sub-lineages of B.1.1.7 (e.g., Q.3) were consolidated into the Alpha category. Baseline surveillance data prior to December 20, 2020 was too thinly sampled to reliably estimate the proportion of viruses from different lineages from this time period, so a second figure was generated using all available sequence data. Since the proportions of Pango lineages over time proved to be very similar in the baseline data and the complete dataset, the larger dataset that dated back to October 2020 was used in the final figure."

Results section from line 91 :
14 genomes were sequenced. It is unclear how they were distributed over the sites, and the number

of samples that would allow sequencing based on Ct values – assuming a cut off of Ct 30 in the E gene PCR, is close to 50. Why were not more samples sequenced? Surely this would have given a much deeper picture. For instance, one could hypothesis that the process of the study itself may have triggered an outbreak on sites 1 and 6 given the baiting involved in getting access to the animals, and the increased prevalence at the second sampling time point.

We have added the site identifiers to Extended Data Table 4. After samples tested positive by the WHO primer sets at OSU, they were retested by the CDC primer sets. Any samples which tested positive by either CDC primer set (N1 or N2) at or below Ct of ~30 at OSU (plus the 2 lowest Ct samples from the March 4 collection date), the original sample material for those samples were sent to NVSL for additional testing and characterization. These 76 samples represented 17 out of the 18 collection dates (on one of the collection dates, March 1, there were no PCR detections by any of the assays run at OSU). Sanger sequencing was initiated on a subset of the positive representing 9 collection dates and all collection reservations. For 7 of those samples, partial sequence was obtained for the spike, nucleocapsid, and RdRp genes with >99% identity to available SARS-CoV-2; for 1 of those samples, partial sequence was obtained for the spike gene with >99% identity to available SARS-CoV-2; for 1 of those samples partial sequencing was unsuccessful. NVSL attempted WGS on all PCR positive samples submitted to NVSL. OSU attempted WGS on all samples with Ct <35 on either CDC primer set (N1 or N2).

Line 104: this is where the relatively sparse coverage of sequencing really is too bad, because the question is whether this is simply explained by undersampling and possibly more recent introduction of the other lineages.

We agree and are actively sampling more deer with the aim of generating more sequence data to address these important questions in the future.

Reviewer Reports on the First Revision:

Two minor additional questions for clarification/revision
a) The authors have addressed the precautions taken to prevent cross contamination or in the RT-PCR process. Can the authors further clarify whether gloves were changed between between each specimen handled?

b) The authors now explain how the deer were baited. Since humans were carrying out the baiting, could infected "baiters" be the source of infection to deer? This does not invalidate the basic findings of the manuscript but may provide a plausible pathway for the spillback infection to occur? It would be useful of the authors comment on this in the discussion.