nature portfolio

Peer Review File

Dominance of Recombinant DWV Genomes with Changing Viral Landscapes as Revealed in National US Honey Bee and Varroa Mite Survey

Corresponding Author: Professor Declan Schroeder

This file contains all reviewer reports in order by version, followed by all author rebuttals in order by version.

Version 0:

Reviewer comments:

Reviewer #1

(Remarks to the Author)
The study reports a collection of viral genomes from bees and Varroa mites collected from various commercial apiaries across the USA, covering 11 states. The authors obtained 383 viromes from bees and 173 from mites. In addition to describing the viral landscape with the frequency of the different viruses found, an important issue related to the diagnosis of DWV variants is highlighted. The results demonstrate the need to monitor the diversity of DWV variants through whole-genome sequencing to generate accurate information. The comparison of the sequences of the viruses found with those of primers commonly used in the literature also underscores the importance of sequencing-based monitoring, rather than relying solely on PCR.
The manuscript reveals a tremendous effort carried out not only in sampling but also in obtaining the genomes and their subsequent analysis. In my opinion, the writing is clear, with well-stated objectives and perfectly described results. However, some questions and comments arise that I detail below:
1. Did any of the sampled colonies show signs of viral infection? Were all virus detections asymptomatic?
2. Did you have information regarding the management of the apiaries that could be used to associate with the presence of viruses? For example: sedentary or migratory management, colony strength, varroa treatment, presence of other pathogens…
3. Was there any colony loss in any of the apiaries?
4. Were you able to confirm the detected genomes by RT-qPCR or RT-PCR in any case?
5. In the sampled apiaries, is Varroa controlled? If so, were the samples taken before or after the treatment?
6. Do you think that, given the objective was to obtain RNA virus genomes, the yield would have improved if an RNA genome enrichment had been performed?
7. In migratory apiaries, the presence and abundance of viruses may vary according to the site to which they migrate, that is, they may acquire new viruses when migrating and return with them to the place of origin. Perhaps it could be indicated, at least in supplementary material, the sampling date and whether it was before or after the movement of the colonies.
8. In the design, a sample was taken from one colony per apiary. Is there intercolony/intra-apiary variation in virus diversity? That is, is a single sample representative of the viral diversity at the apiary level?
9. Were other less frequent viruses found, as described in other studies?

Minor details:
- Page 76: I suggest referring to the Supplementary Material where the number of colonies per apiary and per zone is reported.
- Page 731: In the legend of Figure ExtFig3, change the first (d) to (b).
- Pages 237-239: The authors mention that time series metagenomics will reveal the impact of seasonal variation on the prevalence of type A. Do you think that in a single season the variation in DWV genomes is large enough to make one variant prevail over another?

Reviewer #2

(Remarks to the Author)

Review of Manuscript COMMSBIO-24-4105-T by Benjamin Lamp
“Dominance of Recombinant DWV Genomes with Changing Viral Landscapes as Revealed in Two National US Honey Bee and Varroa Mite Surveys”
Key Messages:
This study sequenced samples from the Bee Informed Partnership's bee health monitoring program. The viromes of commercially kept bees and their ectoparasitic mites were analyzed at the strain level using metagenomics. Viral sequences from 383 adult bees and 173 mites across 11 US states were identified, yielding 45 complete and 1,702 partial genomes. Notably, 312 Sinaivirus, 280 Sacbrood virus, and 135 DWV genomes were deposited in databases. Recombinant DWV genomes, comprising fragments from types A and B, were prevalent. The study raises concerns about the potential overestimation of type A prevalence due to current diagnostic assays, which may not distinguish between recombinant and type A genomes effectively. The nucleotide divergence in SBV and ABPV also causes false negatives in diagnostics. Overall, the study is methodologically sound, well-evaluated, and offers significant new insights. The finding that no major correlations exist between geographical origin and virus composition, likely due to the nationwide movement of bee colonies for pollination, is epidemiologically intriguing and supports the study’s conclusions. While the manuscript is well-written and comprehensive, addressing the following points would enhance its value:
Major Points for Consideration:
1. The use of a template-switching RT enzyme mix, as mentioned in the Methods (Line 302), may have led to an overestimation of recombinant genomes. This enzyme can add non-templated nucleotides to the 3'-end of cDNA, potentially generating artificial recombinants during hybridization with matching RNA genomes. The presence of both DWV-A and DWV-B variants in addition to recombinants should be verified, and this technical issue should be clearly acknowledged and discussed.
Minor Points for Improvement:
1. Line 147: DWV has a monocistronic genome encoding a polyprotein in a single ORF. Thus, there are no 'intergenic regions' in the genome. Please clarify this point, possibly referring to “processing sites between mature proteins.”
2. Previous studies, including ours, have shown that despite their error-prone polymerase, DWV-A and DWV-B variants have remained stable over decades, with only minor mutations observed. This new study contributes a vast number of full-genome sequences, expanding our understanding of DWV diversity. It would be valuable to present a multialignment of this primary dataset at the nucleotide level, alongside reference genomes, to create a master sequence for DWV-B and clarify the differences between these variants. Our group was able to present functionally tested genomes that were established in reverse genetic systems. As these are recent isolates, a comparison of the variants from the USA might be useful to draw epidemiological conclusions that include Europe. Of course, this does not mean that our studies must be cited!
3. DWV-C is notably absent from the discussion. While our group has never detected this variant, it has been reported with high prevalence in the UK and USA. It is possible that our assays were not optimized for DWV-C detection. Was DWV-C detected using metagenomics in this study? If not, why? Clarifying these findings and their implications would strengthen the manuscript.

Reviewer #3

(Remarks to the Author)
Data on virus sequence dynamics is important to understand their impact on hosts and the development of predictive models of virus-host interactions. Invertebrate RNA viruses are poorly characterized compared to those infecting human and other mammals. Therefore, results of this study, which investigated viromes in honey bees and Varroa mites, using nanopore high-throughput sequencing in several hundreds of colonies across the USA, are very valuable.

The major novel finding of this study is that recombinants between DWV-A and DWV-B are widespread in the USA. Novel types of recombinants (e.g. those with recombinant points in the helices gene) were detected. It should be emphasized, that DWV-B (VDV1) arrived to North America much later than DWV-A (see citation # 23, Ryabov et al 2017). Therefore appearance of the DWV-A/DWV-B recombinants occurred in since early 2010s.
This study made a "snapshot" of dynamic DWV complexes occurring in nature, providing a "real-life" example of virus evolutionary trends following arrival of an evolutionary distinct variant. Considering recent arrival of DWV-B in the USA/Americas it would be important to mention the years of honeybee survey in the Abstract.

The study also convincingly showed differences of diversity of virus complexes in honey bees and Varroa mites which reflects differences between honey bee and Varroa mites in their ability to support replication of different pathogens (viruses). I suggest to include a comparison of microbe compositions in honeybees and Varroa using samples derived from the same colonies and include result of this analysis as an additional panel in Fig. 1.

Important point which is made in this MS is that commonly used (q)RT-PCR virus detection methods (which usually detect only a very small genome sections and/or do not produce sequencing data) do not allow to detect recombinant variants and may miss virus sequences due to mismatches between primers and viral RNA targets (L. 178-191). DWV-A in the USA was shown to have high variability (a paper reporting this could be cited https://doi.org/10.1371/journal.pbio.3000502). This study highlights the importance of wider use of unbiassed high-throughput sequencing methods to characterize virus populations.

I think that this MS contain sufficiently novel results to warrant publication in Communications Biology. I have only minor points which should be addressed:

Suggestions and comments:

Abstract. Indicate years of survey - DWV-B became widespread in the USA since mid 2010s.

L. 80 "we sequenced 71.8 million reads" -> "we generated 71.8 million sequencing reads"

L.83-88. Specify average (+- SD) length of viral reads used for the analysis / virus genome assembly.

L. 147. Intergenic regions" - DWV has a single ORF encoding for a viral polyprotein which is cleaved to individual proteins. There are 5'- and 3'-terminal non-coding regions, but no "intergenic regions".

Table 3.
Extend the table to include data on the sequence type for the 5' untranslated region and L-protein.
Indicate nucleotide positions of the sections "Vp1", Vp-2, Vp-3, Helicase, RdRp. What is the "intergenic region"? If it is a region between Halicase and RdRp, this sections contains Vpg and Protease domains.

Fig. 3a.
Modify representation of DWV genomic RNA, indicate position of ORF, 5', 3' ends. Describe what Vp1 - 3, H., RdRp means. I suggest to modify schematic representation of DWV genome according to Ref. 84, (Lanzi et al 2006, Fig 3 A).

L.145 / Table 3. "...thirteen distinct variants of recombinants [DWV] genomes ...". Actual nucleotide sequences of these 13 novel recombinants should be included as a Supplementary text to allow independent analysis.

Version 1:

Reviewer comments:

Reviewer #2

(Remarks to the Author)
Dear authors,

Thank you very much for the detailed answers that clarified my questions from the first review round. Even if I cannot follow your explanations in all points (for example, SMART-RACE reactions tend to terminate prematurely and require a large number of clones to actually determine the 5'-end of an RNA), I have no doubt about the validity of your investigations. I also have no further requests for changes and have read the study with great interest.

Best regards,

Benjamin Lamp

Response to reviewers

Referee expertise:

Referee #1: Biology of honey production

Referee #2: Agricultural virology

Referee #3: Virus-invertebrate host interactions

Reviewers' comments:

Reviewer #1 (Remarks to the Author):

The study reports a collection of viral genomes from bees and Varroa mites collected from various commercial apiaries across the USA, covering 11 states. The authors obtained 383 viromes from bees and 173 from mites. In addition to describing the viral landscape with the frequency of the different viruses found, an important issue related to the diagnosis of DWV variants is highlighted. The results demonstrate the need to monitor the diversity of DWV variants through whole-genome sequencing to generate accurate information. The comparison of the sequences of the viruses found with those of primers commonly used in the literature also underscores the importance of sequencing-based monitoring, rather than relying solely on PCR.

The manuscript reveals a tremendous effort carried out not only in sampling but also in obtaining the genomes and their subsequent analysis. In my opinion, the writing is clear, with well-stated objectives and perfectly described results.

Response: Thank you!

However, some questions and comments arise that I detail below:
1. Did any of the sampled colonies show signs of viral infection? Were all virus detections asymptomatic?

Response: The colonies sampled for this study were selected at random as part of a larger monitoring program designed for commercial beekeepers, which provides four inspection visits per year at critical moments of the season. Because the colonies were chosen at random within each operation, the specific symptomatic or asymptomatic status of individual colonies at the time of sampling was not a selection criterion. The aim was to ensure that the sampled colonies were representative of the overall state of the apiary, without introducing bias toward either "healthy" or "unhealthy" colonies. Field technicians are trained to sample without bias and not to select colonies based on perceived health status.

It is important to note that many viral infections in honey bees may not exhibit clear or consistent symptoms. In many cases, symptomatic signs of viral infections are non-specific or can overlap with other stressors such as pesticide exposure, bacterial infection, etc. Therefore, diagnosing viral infections based solely on visual inspection

is not reliable. In addition, there is a growing consensus in the field that viral infections often have an asymptomatic phase, which adds to the complexity of associating viral presence with colony health outcomes. This is one reason why our study emphasises the importance of genomic monitoring over reliance on symptomatic observation alone.
That said, this remains an important point and as such we changed the first sentence in M&M (line 305) to read as follows: "*A random subsample of honey bees and mites were collected from around the US from commercial beekeepers in the US that follow migratory practices. These beekeepers are part of the larger cohort of apiaries that are routinely monitored by the Bee Informed Partnership. The aim was to ensure that the sampled colonies were representative of the overall state of the apiary, without introducing bias toward either healthy or unhealthy colonies*.".**

2. Did you have information regarding the management of the apiaries that could be used to associate with the presence of viruses? For example: sedentary or migratory management, colony strength, varroa treatment, presence of other pathogens…

Response: Thank you for this question. All beekeepers participating in this study are part of the commercial sector, and nearly all commercial beekeepers in the USA follow migratory practices. However, gathering detailed management data from these operations is notoriously challenging, and organising it into analyzable groups is even more difficult.

Commercial beekeepers typically manage their colonies at the operation level, rather than focusing on individual colonies. For instance, it is common for beekeepers to use various migratory routes and schedules, often sending different batches of colonies to different locations without tracking the individual movements of each colony. Colonies are all broken down in the spring when creating new splits, and most often combined in the fall. Treatments and feeding are usually done "as needed," often with homemade recipes and customised application methods, making it nearly impossible to standardise these practices for comparative analysis. As a result, it is rare to follow the same colony over time or track its individual history or fate.

We do have access to some management-related covariates, such as Varroa mite loads, and the record of signs of foulbrood or fungal infections at the time of sampling. However, attempting to analyse associations between these variables and viral presence without a clear, hypothesis-driven approach could easily lead to spurious correlations, given the high variability in management practices across operations.

Therefore and in keeping with point 1 above, we modified the first sentence in M&M (line 305) to read as follows: "*A random subsample of honey bees and mites were collected from around the US from commercial beekeepers in the US that follow migratory practices. These beekeepers are part of the larger cohort of apiaries that are routinely monitored by the Bee Informed Partnership.* The aim was to ensure that the sampled colonies were representative of the overall state of the apiary, without introducing bias toward either healthy or unhealthy colonies."

3. Was there any colony loss in any of the apiaries?

Response: We agree that it would be highly valuable to relate viral loads to colony outcomes, particularly winter survival or loss. However, this study was not designed to achieve that specific goal.

Tracking colony outcomes would require participating beekeepers to isolate sampled colonies from their typical fall management practices, including preventing them from combining colonies. It would also necessitate tracking those colonies as they are relocated, often to different locations after winter, and then reporting their fate back to us. Unfortunately, most commercial beekeepers are not able or willing to make these accommodations due to the operational challenges it poses.

Although this was not a primary objective of our study, we did attempt to collect operational-level data on winter colony losses. However, we received limited responses from the participating beekeepers, which hindered our ability to draw conclusions on this aspect.

4. Were you able to confirm the detected genomes by RT-qPCR or RT-PCR in any case?

Response: This was beyond the scope of this study, however, please refer to our prior work where we have demonstrated that RT-qPCR data concurs with ONT sequencing data generated via our published method (Schroeder Lab), especially highlighting the detection of a recombinant genomes of DWV that were missed by RT-qPCR (Nikulin et al 2024). Also, as stated above, these commercial beekeepers are part of the larger Bee Informed Partnership monitoring program that uses RT-qPCR as the main tool for virus surveillance. We discuss this in our manuscript, highlighting that though we broadly find similar results, the resolution and sensitivity of the RT-qPCR method is not as good as the sequencing data.

5. In the sampled apiaries, is Varroa controlled? If so, were the samples taken before or after the treatment?

Response: Almost all commercial beekeepers in the US use chemical treatments to manage Varroa mites, typically applying these treatments multiple times throughout the year. However, the specific techniques, formulations, and timing of treatments vary significantly between operations.

The Bee Informed Partnership monitoring program includes two fall visits, during which they aim to sample colonies both before and after treatment. This is intended to provide the participating beekeepers with an opportunity to assess the effectiveness of their Varroa control strategies. However, it is important to note that the timing of our visits does not always align perfectly with the beekeepers' treatment schedules. As a result, while we can confirm that Varroa was controlled in these operations, we cannot definitively state how close the sample collection occurred to the treatments in every case. In summary, while Varroa treatments were in place, the exact timing of our sampling in relation to the treatment may vary.

6. Do you think that, given the objective was to obtain RNA virus genomes, the yield would have improved if an RNA genome enrichment had been performed?

Response: This is possible but it will introduce other biases such as disproportional quasispecies enrichment. Based on previous work done (e.g. Nikulin et al 2024) we are confident that the genomes generated are indicative of the dominant circulating genomes.

7. In migratory apiaries, the presence and abundance of viruses may vary according to the site to which they migrate, that is, they may acquire new viruses when migrating and return with them to the place of origin. Perhaps it could be indicated, at least in supplementary material, the sampling date and whether it was before or after the movement of the colonies.

Response: Thank you for raising this important point. We fully agree that migratory beekeeping practices can play a significant role in virus transmission, and the potential for colonies to acquire new viruses during migration and bring them back to their place of origin is an area of particular interest to us. In fact, we had proposed a comparative cohort study to explore this topic in more detail, but unfortunately, it was not funded.

For the current study, our aim was to represent commercial operations across major hubs of beekeeping in the US, without focusing on specific migratory routes or events. As such, we did not capture detailed information on the timing of migrations relative to sample collection. Tracking migratory movements at the individual colony level is logistically challenging, as many commercial beekeepers participate in multiple migratory events and often send different subsets of colonies to various locations (without tracking individual colony ids). This variability made it difficult to incorporate such detailed migration data into our study.

While we acknowledge the value of providing the sampling dates and details on whether it occurred before or after migration, this information was not systematically recorded for individual colonies. However, this remains an interesting avenue for future research, and we appreciate your suggestion.

8. In the design, a sample was taken from one colony per apiary. Is there intercolony/intra-apiary variation in virus diversity? That is, is a single sample representative of the viral diversity at the apiary level?

Response: Thank you for raising this important question. The potential for intercolony and intra-apiary variation in viral diversity was indeed a key consideration in our study design.

In our standard monitoring efforts, we typically sample 8 colonies per location or "drop," as most commercial beekeepers manage between 30 and 90 colonies in a single location (outside of major pollination events). These colonies are chosen at random, without intentionally selecting for either visibly healthy or sick colonies.

However, due to the cost associated with molecular analyzes, we opted to analyze only a single colony per apiary in this study. This decision was made to maximise the number of operations represented in our dataset, prioritising breadth of coverage across different apiaries rather than depth within each operation.

Previous studies suggest that operation-to-operation variability (based on factors like Varroa loads, frames of bees, and other inspection metrics) is generally greater than colony-to-colony variability within the same operation. While this preliminary evidence supports our approach, we acknowledge that this pattern may not fully extend to viral diversity, and further research is needed to confirm these findings in the context of viromes.

In summary, while a single sample may not capture the full extent of viral diversity within an apiary, our design aimed to capture broader trends across operations, which we believed was most important for the goals of this study.

9. Were other less frequent viruses found, as described in other studies?

Response: As we have not taken the same approach as others, it would be inappropriate to directly compare our results with these other studies. That said, it is clear that we do pick up the usual suspects (DWV, SBV & LSV). What was unique was the level of DWV recombinants found, which is the main message of the paper. And that other studies would not have detected these because of the RT-PCR approach taken.

Minor details:
- Page 76: I suggest referring to the Supplementary Material where the number of colonies per apiary and per zone is reported.

Response: Done (lines 81)

- Page 731: In the legend of Figure ExtFig3, change the first (d) to (b).

Response: changed (d) to (b)

- Pages 237-239: The authors mention that time series metagenomics will reveal the impact of seasonal variation on the prevalence of type A. Do you think that in a single season the variation in DWV genomes is large enough to make one variant prevail over another?

Response: We preface this statement by saying (line 243-244): "*Past work has indicated that seasons have no bearing on DWV load or prevalence in colonies in the United Kingdom*[12],". So we are not stating that season will be impactful, but who knows… these viruses do change fast and especially if recombination is the main mechanism underpinning this change. And a single season could be long enough to observe these changes, as stated in lines 244-246.**

Reviewer #2 (Remarks to the Author):

Review of Manuscript COMMSBIO-24-4105-T by Benjamin Lamp

“Dominance of Recombinant DWV Genomes with Changing Viral Landscapes as Revealed in Two National US Honey Bee and Varroa Mite Surveys”
Key Messages:

This study sequenced samples from the Bee Informed Partnership's bee health monitoring program. The viromes of commercially kept bees and their ectoparasitic mites were analyzed at the strain level using metagenomics. Viral sequences from 383 adult bees and 173 mites across 11 US states were identified, yielding 45 complete and 1,702 partial genomes. Notably, 312 Sinaivirus, 280 Sacbrood virus, and 135 DWV genomes were deposited in databases. Recombinant DWV genomes, comprising fragments from types A and B, were prevalent. The study raises concerns about the potential overestimation of type A prevalence due to current diagnostic assays, which may not distinguish between recombinant and type A genomes effectively. The nucleotide divergence in SBV and ABPV also causes false negatives in diagnostics. Overall, the study is methodologically sound, well-evaluated, and offers significant new insights. The finding that no major correlations exist between geographical origin and virus composition, likely due to the nationwide movement of bee colonies for pollination, is epidemiologically intriguing and supports the study’s conclusions. While the manuscript is well-written and comprehensive, addressing the following points would enhance its value:

Major Points for Consideration:

1. The use of a template-switching RT enzyme mix, as mentioned in the Methods (Line 302), may have led to an overestimation of recombinant genomes. This enzyme can add non-templated nucleotides to the 3'-end of cDNA, potentially generating artificial recombinants during hybridization with matching RNA genomes. The presence of both DWV-A and DWV-B variants in addition to recombinants should be verified, and this technical issue should be clearly acknowledged and discussed.

Response: The reviewer is correct to point out that we need to be certain that the method used do not create unwanted diversity such as recombinants. The template switching RT enzyme works through an addition of three cytosines when it reaches the 5’ end of the mRNA/viral genome molecule. Firstly, this enzyme does not readily fall off until it gets to the 5’ end. This is why it was used extensively in 5’ RACE experiments (SMARTer technology by Takara: https://www.takarabio.com/learning-centers/next-generation-sequencing/technology-and-application-overviews/smart-technology?srsltid=AfmBOoo0TYqpoeOwSZHOh_eTPK1zBZflVXiGA0oowu5ywEp_W281zkQa). Secondly, a template switching oligo that binds (hybridises) to the three cytosines is absolutely required (so it's targeted) and this is what is used for template switching. So it’s not a random switching event based on random genome hybridizations.

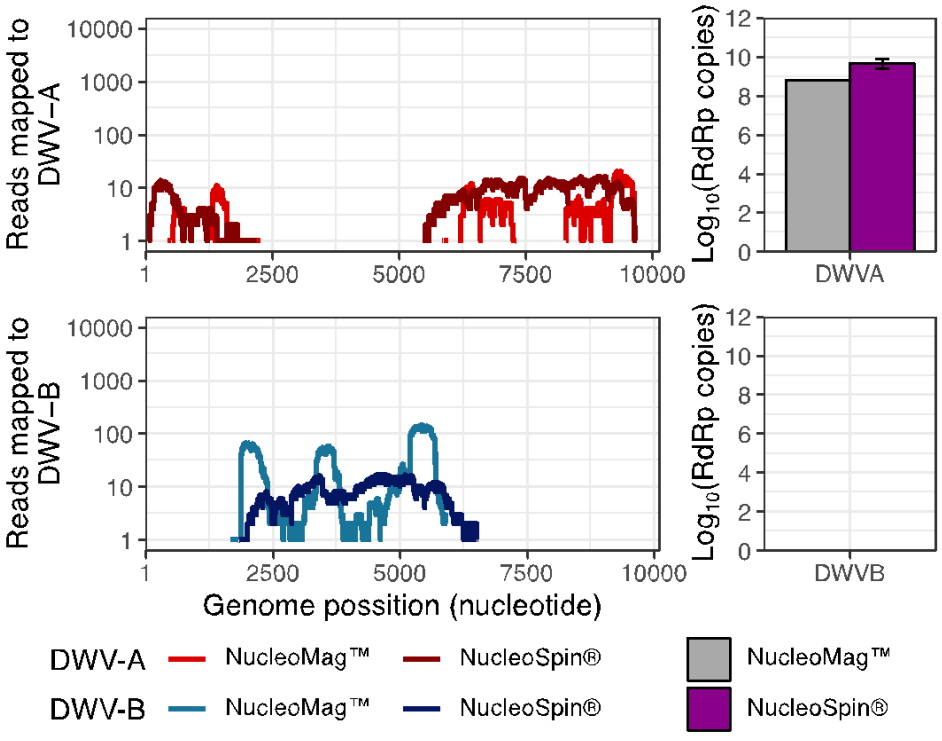
It must be remembered that template switching–based library preparation has emerged as a method of choice for single cell sequencing because of its relatively simple workflow and low RNA input requirements (Ziegenhain et al. 2017). It has also been adapted to sequencing small RNAs as well as fragmented RNAs from liquid

biopsies or microsomes (Rantalainen, 2018). Furthermore, template switching has been used in long-read single-molecule approaches with third-generation sequencing platforms, including isoform (Iso)-Seq (PacBio SMART-Seq) (Cartolano et al 2016) and nanopore RNA-Seq (Oxford Nanopore Technology) (Oikonomopoulos et al. 2016).

The reviewer should also be aware that we (Schroeder Lab) used this technology in other host-virus systems, e.g. sequenced PRRSV in pigs (Schroeder et al 2021) and MeV in humans (Yousaf et al., 2023). The Yousaf study used the Schroeder Lab's template switching sequencing pipeline to validate their Illumina based assemblies. NO recombinants were reported in this study.

Similarly, the LSV genomes that were recovered in the honey bees screen in Guyana did not have recombinants either (Schroeder et al 2022). And in our honey bee viromes associated with EFB (Hesketh-Best et al 2024), we did not recover many DWV recombinants, most were wild-type A or B (figure 2).

Finally, our entire sequencing workflow was further validated in the Nikulin et 2024 study. Here we demonstrated that RT-qPCR data concurs with ONT sequencing data generated where we confirmed the presence of recombinant genomes of DWV that were missed by qPCR detection (Figure S1 - see below for a copy). For this work we mapped reads against the same reference genomes for DWV-A and B genomes, as used in this study. We observed coverage of DWV-A that drops drastically before observing an increase in B, which then returns to a high coverage of DWV-A. We confirmed this region within the subsequent contigs generated from this sample and confirmed breakpoints (HyPhy-GARD), in a similar manner as performed within the current manuscript.



S1 Fig. Both extraction protocols are capable of identifying chimeric DWV genomes.
Colony # 8, Location: Hill, Year: 2021. Coverage histograms (left) represented as the number of reads mapped to two Deformed wing virus' strains from NCBI, (top) DWV-A (NC_004830.2) and (bottom) DWV-B/VDV-1 (NC_006494.1). Copy number of RNA-dependent RNA polymerase for the two DWV strains as detected by RT-qPCR (right).

Furthermore, for a random subset of the genomes generated in the current manuscript, due to the high labour intensity of this confirmation, we did observe multiple long-reads spanning the breakpoints identified within this present study - not presented within the manuscript due to brevity.

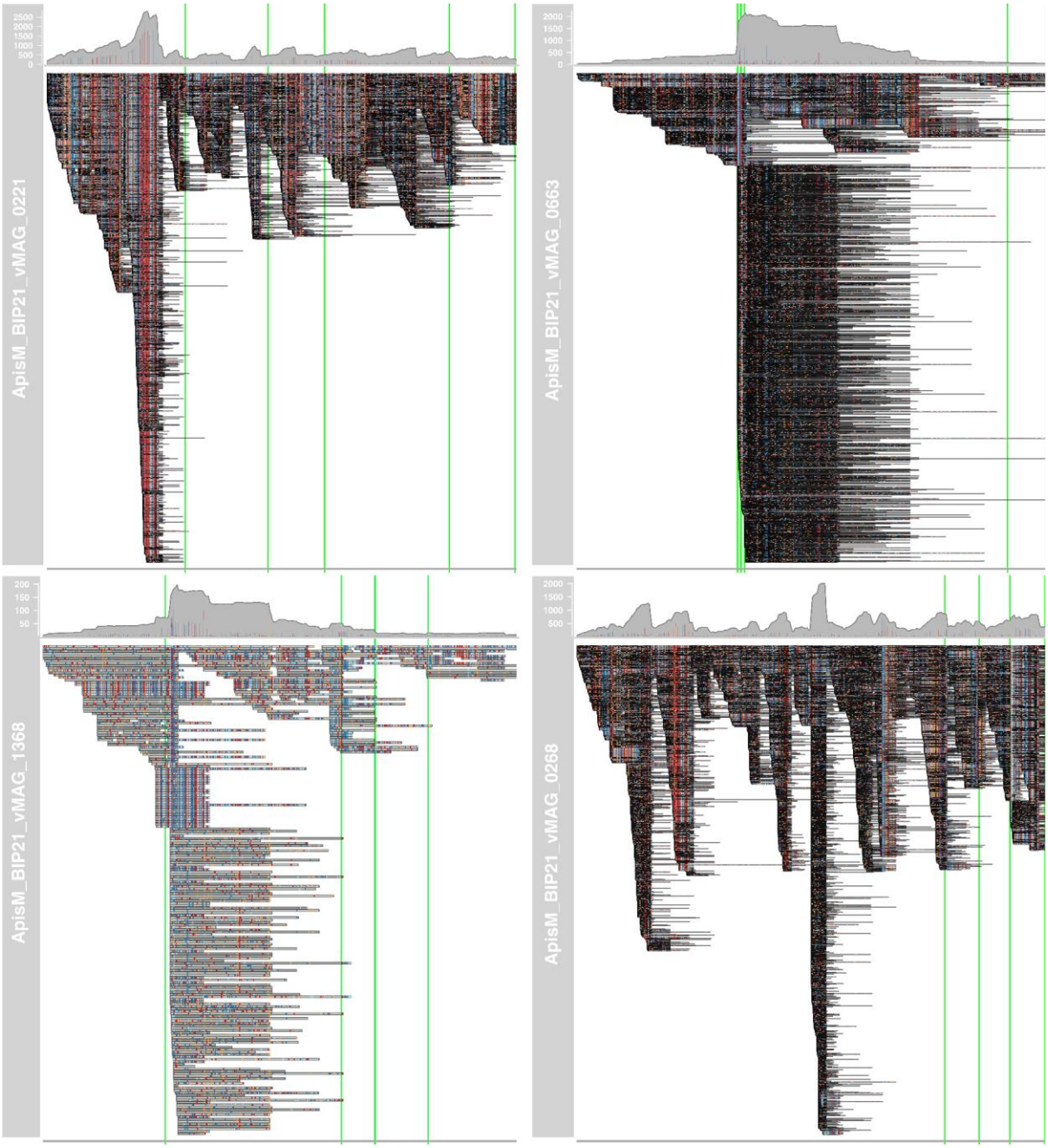


Figure. Validation of recombination breakpoint through the inspection of read mapping and coverage. We observe breakpoints as having multiple reads spanning 50 bp on either side of the breakpoint for recombinant genomes assembled in this study. Coverage (top) and read (bottom) plots of four recombinant genomes are presented here, with the HyPhy-GARD predicted breakpoint locations denoted as green lines.

Our recombinant hotspots are in agreement with prior literature (largely in the '5 UTR, Helicase (Dalmon et al 2017)). Discussed in the manuscript at lines 230-237. The arguments presented in this manuscript and here would suggest that while we can never rule out the possibility of chimeric assemblies, the genomic structures of our recombinant genomes are concurrent with the literature, our manuscript simply presents data that indicates recombinant genomes are more frequent than previously reported.

To conclude, we are confident that the method or technology used is not creating false recombinants.

Minor Points for Improvement:

1. Line 147: DWV has a monocistronic genome encoding a polyprotein in a single ORF. Thus, there are no 'intergenic regions' in the genome. Please clarify this point, possibly referring to "processing sites between mature proteins."

Response: We agree, thank you for the suggestion of this phrase, we have made the change throughout the manuscript.

2. Previous studies, including ours, have shown that despite their error-prone polymerase, DWV-A and DWV-B variants have remained stable over decades, with only minor mutations observed. This new study contributes to a vast number of full-genome sequences, expanding our understanding of DWV diversity. It would be valuable to present a multialignment of this primary dataset at the nucleotide level, alongside reference genomes, to create a master sequence for DWV-B and clarify the differences between these variants. Our group was able to present functionally tested genomes that were established in reverse genetic systems. As these are recent isolates, a comparison of the variants from the USA might be useful to draw epidemiological conclusions that include Europe. Of course, this does not mean that our studies must be cited!

Response: Given the scale of our data set (an example of a smaller data set shown above in major point 1), this type of figure would be impractical for a manuscript. Our sequences are available and such an alignment can easily be created by anyone. We also don't think that this has a practical application as viruses exist as quasi-species. So many of the nucleotide mutations can easily be lost or be part of a defective genome. For this reason we focused our attention on the protein sequence.

3. DWV-C is notably absent from the discussion. While our group has never detected this variant, it has been reported with high prevalence in the UK and USA. It is possible that our assays were not optimized for DWV-C detection. Was DWV-C detected using metagenomics

in this study? If not, why? Clarifying these findings and their implications would strengthen the manuscript.

Response: Correct, DWV-C was reported by us in the US in the Kevill et al 2019 paper. But this genome was not common and when it was found it was mainly in New Mexico. Moreover, type C was first reported in Devon, UK (also by the Schroeder Lab) in 2016 (Mordecai et al. 2016). Notably it was absent in the 2019 UK survey (Kevill et al 2019). So again it's clear that DWV-C is not abundant and not easy to find. It's also likely that honey bees are not its primary host. We found that it was dominant in small hive beetles (Brettell et al 2019). Interestingly, whole genome sequencing using Illumina in this study showed that the DWV-C in small hive beetles was a recombinant. Therefore, it's not unexpected that we did not find DWV-C in our study. Making this point could be a distraction as it is largely based on an hypothesis that we were not testing here.

That said, we have added the following statement to make readers aware of the lower limits of detection of our study, which includes both DWV-C and the recently reported A. mellifera solinvivirus-1 (AmSV1): *"Nor did we identify any DWV type C reads or contigs."* (line 226). We hope this emphasises to readers that not all viruses that have been detected in honey bees viromes were captured within this study, as this paragraph is concluded with the following statement: *"The absence of these uncommon viruses demonstrates a limit in detection of this study, possibly a result of our sequencing efforts not being expansive enough to detect low-titer viruses or our sampling occurred at a period of particularly low density."* (Lines 226-229)

Reviewer #3 (Remarks to the Author):

Data on virus sequence dynamics is important to understand their impact on hosts and the development of predictive models of virus-host interactions. Invertebrate RNA viruses are poorly characterized compared to those infecting human and other mammals. Therefore, results of this study, which investigated viromes in honey bees and Varroa mites, using nanopore high-throughput sequencing in several hundreds of colonies across the USA, are very valuable.

The major novel finding of this study is that recombinants between DWV-A and DWV-B are widespread in the USA. Novel types of recombinants (e.g. those with recombinant points in the helices gene) were detected. It should be emphasized, that DWV-B (VDV1) arrived to North America much later than DWV-A (see citation # 23, Ryabov et al 2017).

Response: Thank you for pointing this out. We modified the sentence in line 64 to emphasise this point: "Type B arrived much later in the US than type A[17], and until recently, type A was the most common variant in the US and potentially worldwide[12,23]" (citing Ryabov et a. 2017 buch earlier - now citation #17.

Therefore appearance of the DWV-A/DWV-B recombinants occurred in since early 2010s. This study made a "snapshot" of dynamic DWV complexes occurring in nature, providing a "real-life" example of virus evolutionary trends following arrival of an evolutionary distinct

variant. Considering recent arrival of DWV-B in the USA/Americas it would be important to mention the years of honeybee survey in the Abstract.

The study also convincingly showed differences of diversity of virus complexes in honey bees and Varroa mites which reflects differences between honey bee and Varroa mites in their ability to support replication of different pathogens (viruses).

Response: Thank you for this summary and positive feedback

I suggest to include a comparison of microbe compositions in honeybees and Varroa using samples derived from the same colonies and include result of this analysis as an additional panel in Fig. 1.

Response: Correct, we do have the microbial data BUT this is more complicated to present in the context of this study. This is because this is metatranscriptomics data and not the typical barcoding type 16S datasets. Moreover, the host genomes (nuclear and mitochondrial) also make up a significant proportion of the transcriptome. For this reason we chose not to include this analysis in this manuscript. This will be the subject of another paper.

Important point which is made in this MS is that commonly used (q)RT-PCR virus detection methods (which usually detect only a very small genome sections and/or do not produce sequencing data) do not allow to detect recombinant variants and may miss virus sequences due to mismatches between primers and viral RNA targets (L. 178-191). DWV-A in the USA was shown to have high variability (a paper reporting this could be cited https://doi.org/10.1371/journal.pbio.3000502). This study highlights the importance of wider use of unbiassed high-throughput sequencing methods to characterize virus populations.

Response: Agreed, we inserted the following sentence, "Furthermore, DWV-A has been observed to have high genetic variability, which may further complicate surveillance efforts[39]." (line 280), with citation.

I think that this MS contain sufficiently novel results to warrant publication in Communications Biology. I have only minor points which should be addressed:

Suggestions and comments:

Abstract. Indicate years of survey - DWV-B became widespread in the USA since mid 2010s.
Response: Done (Line 26)

L. 80 "we sequenced 71.8 million reads" -> "we generated 71.8 million sequencing reads"
Response: Done (Line 81)

L.83-88. Specify average (+- SD) length of viral reads used for the analysis / virus genome assembly.

Response: **The average read length of the sequence data has now been stated (mean= 415 bp +/- 124 bp) - Line 83.**

L. 147. Intergenic regions" - DWV has a single ORF encoding for a viral polyprotein which is cleaved to individual proteins. There are 5'- and 3'-terminal non-coding regions, but no "intergenic regions".

Response: **We have changed this to the region 'processing sites between mature proteins.' as per the suggestion of reviewer 2. Including in Table 3.**

Table 3.
Extend the table to include data on the sequence type for the 5' untranslated region and L-protein.

Response: **Assembly of a complete 5' UTR (and non contains complete 3' terminal ends of the genome) was uncommon. Table 3 reports the structure of only genomes with complete regions (excluding partial regions of the genomes). We have added the statement: "*The assembly of the 5' UTR was uncommon, and we did not obtain any DWV genomes with complete 3'-terminal non-coding regions.*" lines 151-152.**

To the table 3 legend we also added the following: "*We excluded the untranslated regions (UTR) of DWV from the following analysis and designated all genomic regions that are not mature protein sequences as 'Regions between mature proteins'*". See response to the comment below as to why L-proteins are not reported.

Indicate nucleotide positions of the sections "Vp1", Vp-2, Vp-3, Helicase, RdRp. What is the "intergenic region"? If it is a region between Halicase and RdRp, this section contains Vpg and Protease domains.

Response: **Due to the high number of genomes and to ensure replicability of these methods, we utilised InterProScan, which did not functionally classify the protease domains and L-proteins. Therefore we don't report the exact positional data.**

Fig. 3a.
Modify representation of DWV genomic RNA, indicate position of ORF, 5', 3' ends. Describe what Vp1 - 3, H., RdRp means. I suggest to modify schematic representation of DWV genome according to Ref. 84, (Lanzi et al 2006, Fig 3 A).

Response: **As stated above, these annotations are made utilising InterProScan, a tool that combines different protein signature recognition methods into a single output (see MM subsection '*Genome functional annotation and pSSN*'), and in the case of this study we utilised 6 distinct protein databases. We are disinclined to incorporate this suggestion as we do not feel a method based on the estimation of protein annotations using a single publication is a scalable or reproducible approach for metagenomic studies.**

This is indeed a limitation of our annotation approach, and as such we have added the following statement to make readers aware that not all known genomic genomes of

DWV are annotated: "*Our annotation method was limited by the fact that DWV L-proteins and proteases were not identified by InterProScan (#27, Lanzi et al 2006). Consequently, we are unable to comment on the frequency or occurrence of recombination within these genomic regions*" (lines 161-164).

L.145 / Table 3. "...thirteen distinct variants of recombinants [DWV] genomes ...". Actual nucleotide sequences of these 13 novel recombinants should be included as a Supplementary text to allow independent analysis.

Response: All full fasta file sequences of these contigs are present in the data repository, as stated in the Data and code availability statement. We feel that we included sufficient metadata to aid researchers to partition the recombinant genomes. Therefore we feel that an additional supplementary data file just for thirteen genomes is unnecessary in a manuscript already contains a large number of supplemental data files.

Response to reviewers

Referee expertise:

Referee #1: Biology of honey production

Referee #2: Agricultural virology

Referee #3: Virus-invertebrate host interactions

Reviewers' comments:

Reviewer #1 (Remarks to the Author):

The study reports a collection of viral genomes from bees and Varroa mites collected from various commercial apiaries across the USA, covering 11 states. The authors obtained 383 viromes from bees and 173 from mites. In addition to describing the viral landscape with the frequency of the different viruses found, an important issue related to the diagnosis of DWV variants is highlighted. The results demonstrate the need to monitor the diversity of DWV variants through whole-genome sequencing to generate accurate information. The comparison of the sequences of the viruses found with those of primers commonly used in the literature also underscores the importance of sequencing-based monitoring, rather than relying solely on PCR.

The manuscript reveals a tremendous effort carried out not only in sampling but also in obtaining the genomes and their subsequent analysis. In my opinion, the writing is clear, with well-stated objectives and perfectly described results.

Response: Thank you!

However, some questions and comments arise that I detail below:
1. Did any of the sampled colonies show signs of viral infection? Were all virus detections asymptomatic?

Response: The colonies sampled for this study were selected at random as part of a larger monitoring program designed for commercial beekeepers, which provides four inspection visits per year at critical moments of the season. Because the colonies were chosen at random within each operation, the specific symptomatic or asymptomatic status of individual colonies at the time of sampling was not a selection criterion. The aim was to ensure that the sampled colonies were representative of the overall state of the apiary, without introducing bias toward either "healthy" or "unhealthy" colonies. Field technicians are trained to sample without bias and not to select colonies based on perceived health status.

It is important to note that many viral infections in honey bees may not exhibit clear or consistent symptoms. In many cases, symptomatic signs of viral infections are non-specific or can overlap with other stressors such as pesticide exposure, bacterial infection, etc. Therefore, diagnosing viral infections based solely on visual inspection

is not reliable. In addition, there is a growing consensus in the field that viral infections often have an asymptomatic phase, which adds to the complexity of associating viral presence with colony health outcomes. This is one reason why our study emphasises the importance of genomic monitoring over reliance on symptomatic observation alone.
That said, this remains an important point and as such we changed the first sentence in M&M (line 305) to read as follows: "*A random subsample of honey bees and mites were collected from around the US from commercial beekeepers in the US that follow migratory practices. These beekeepers are part of the larger cohort of apiaries that are routinely monitored by the Bee Informed Partnership. The aim was to ensure that the sampled colonies were representative of the overall state of the apiary, without introducing bias toward either healthy or unhealthy colonies*.".**

2. Did you have information regarding the management of the apiaries that could be used to associate with the presence of viruses? For example: sedentary or migratory management, colony strength, varroa treatment, presence of other pathogens…

Response: Thank you for this question. All beekeepers participating in this study are part of the commercial sector, and nearly all commercial beekeepers in the USA follow migratory practices. However, gathering detailed management data from these operations is notoriously challenging, and organising it into analyzable groups is even more difficult.

Commercial beekeepers typically manage their colonies at the operation level, rather than focusing on individual colonies. For instance, it is common for beekeepers to use various migratory routes and schedules, often sending different batches of colonies to different locations without tracking the individual movements of each colony. Colonies are all broken down in the spring when creating new splits, and most often combined in the fall. Treatments and feeding are usually done "as needed," often with homemade recipes and customised application methods, making it nearly impossible to standardise these practices for comparative analysis. As a result, it is rare to follow the same colony over time or track its individual history or fate.

We do have access to some management-related covariates, such as Varroa mite loads, and the record of signs of foulbrood or fungal infections at the time of sampling. However, attempting to analyse associations between these variables and viral presence without a clear, hypothesis-driven approach could easily lead to spurious correlations, given the high variability in management practices across operations.

Therefore and in keeping with point 1 above, we modified the first sentence in M&M (line 305) to read as follows: "*A random subsample of honey bees and mites were collected from around the US from commercial beekeepers in the US that follow migratory practices. These beekeepers are part of the larger cohort of apiaries that are routinely monitored by the Bee Informed Partnership.* The aim was to ensure that the sampled colonies were representative of the overall state of the apiary, without introducing bias toward either healthy or unhealthy colonies."

3. Was there any colony loss in any of the apiaries?

Response: We agree that it would be highly valuable to relate viral loads to colony outcomes, particularly winter survival or loss. However, this study was not designed to achieve that specific goal.

Tracking colony outcomes would require participating beekeepers to isolate sampled colonies from their typical fall management practices, including preventing them from combining colonies. It would also necessitate tracking those colonies as they are relocated, often to different locations after winter, and then reporting their fate back to us. Unfortunately, most commercial beekeepers are not able or willing to make these accommodations due to the operational challenges it poses.

Although this was not a primary objective of our study, we did attempt to collect operational-level data on winter colony losses. However, we received limited responses from the participating beekeepers, which hindered our ability to draw conclusions on this aspect.

4. Were you able to confirm the detected genomes by RT-qPCR or RT-PCR in any case?

Response: This was beyond the scope of this study, however, please refer to our prior work where we have demonstrated that RT-qPCR data concurs with ONT sequencing data generated via our published method (Schroeder Lab), especially highlighting the detection of a recombinant genomes of DWV that were missed by RT-qPCR (Nikulin et al 2024). Also, as stated above, these commercial beekeepers are part of the larger Bee Informed Partnership monitoring program that uses RT-qPCR as the main tool for virus surveillance. We discuss this in our manuscript, highlighting that though we broadly find similar results, the resolution and sensitivity of the RT-qPCR method is not as good as the sequencing data.

5. In the sampled apiaries, is Varroa controlled? If so, were the samples taken before or after the treatment?

Response: Almost all commercial beekeepers in the US use chemical treatments to manage Varroa mites, typically applying these treatments multiple times throughout the year. However, the specific techniques, formulations, and timing of treatments vary significantly between operations.

The Bee Informed Partnership monitoring program includes two fall visits, during which they aim to sample colonies both before and after treatment. This is intended to provide the participating beekeepers with an opportunity to assess the effectiveness of their Varroa control strategies. However, it is important to note that the timing of our visits does not always align perfectly with the beekeepers' treatment schedules. As a result, while we can confirm that Varroa was controlled in these operations, we cannot definitively state how close the sample collection occurred to the treatments in every case. In summary, while Varroa treatments were in place, the exact timing of our sampling in relation to the treatment may vary.

6. Do you think that, given the objective was to obtain RNA virus genomes, the yield would have improved if an RNA genome enrichment had been performed?

Response: This is possible but it will introduce other biases such as disproportional quasispecies enrichment. Based on previous work done (e.g. Nikulin et al 2024) we are confident that the genomes generated are indicative of the dominant circulating genomes.

7. In migratory apiaries, the presence and abundance of viruses may vary according to the site to which they migrate, that is, they may acquire new viruses when migrating and return with them to the place of origin. Perhaps it could be indicated, at least in supplementary material, the sampling date and whether it was before or after the movement of the colonies.

Response: Thank you for raising this important point. We fully agree that migratory beekeeping practices can play a significant role in virus transmission, and the potential for colonies to acquire new viruses during migration and bring them back to their place of origin is an area of particular interest to us. In fact, we had proposed a comparative cohort study to explore this topic in more detail, but unfortunately, it was not funded.

For the current study, our aim was to represent commercial operations across major hubs of beekeeping in the US, without focusing on specific migratory routes or events. As such, we did not capture detailed information on the timing of migrations relative to sample collection. Tracking migratory movements at the individual colony level is logistically challenging, as many commercial beekeepers participate in multiple migratory events and often send different subsets of colonies to various locations (without tracking individual colony ids). This variability made it difficult to incorporate such detailed migration data into our study.

While we acknowledge the value of providing the sampling dates and details on whether it occurred before or after migration, this information was not systematically recorded for individual colonies. However, this remains an interesting avenue for future research, and we appreciate your suggestion.

8. In the design, a sample was taken from one colony per apiary. Is there intercolony/intra-apiary variation in virus diversity? That is, is a single sample representative of the viral diversity at the apiary level?

Response: Thank you for raising this important question. The potential for intercolony and intra-apiary variation in viral diversity was indeed a key consideration in our study design.

In our standard monitoring efforts, we typically sample 8 colonies per location or "drop," as most commercial beekeepers manage between 30 and 90 colonies in a single location (outside of major pollination events). These colonies are chosen at random, without intentionally selecting for either visibly healthy or sick colonies.

However, due to the cost associated with molecular analyzes, we opted to analyze only a single colony per apiary in this study. This decision was made to maximise the number of operations represented in our dataset, prioritising breadth of coverage across different apiaries rather than depth within each operation.

Previous studies suggest that operation-to-operation variability (based on factors like Varroa loads, frames of bees, and other inspection metrics) is generally greater than colony-to-colony variability within the same operation. While this preliminary evidence supports our approach, we acknowledge that this pattern may not fully extend to viral diversity, and further research is needed to confirm these findings in the context of viromes.

In summary, while a single sample may not capture the full extent of viral diversity within an apiary, our design aimed to capture broader trends across operations, which we believed was most important for the goals of this study.

9. Were other less frequent viruses found, as described in other studies?

Response: As we have not taken the same approach as others, it would be inappropriate to directly compare our results with these other studies. That said, it is clear that we do pick up the usual suspects (DWV, SBV & LSV). What was unique was the level of DWV recombinants found, which is the main message of the paper. And that other studies would not have detected these because of the RT-PCR approach taken.

Minor details:
- Page 76: I suggest referring to the Supplementary Material where the number of colonies per apiary and per zone is reported.

Response: Done (lines 81)

- Page 731: In the legend of Figure ExtFig3, change the first (d) to (b).

Response: changed (d) to (b)

- Pages 237-239: The authors mention that time series metagenomics will reveal the impact of seasonal variation on the prevalence of type A. Do you think that in a single season the variation in DWV genomes is large enough to make one variant prevail over another?

Response: We preface this statement by saying (line 243-244): "*Past work has indicated that seasons have no bearing on DWV load or prevalence in colonies in the United Kingdom*[12],". So we are not stating that season will be impactful, but who knows… these viruses do change fast and especially if recombination is the main mechanism underpinning this change. And a single season could be long enough to observe these changes, as stated in lines 244-246.**

Reviewer #2 (Remarks to the Author):

Review of Manuscript COMMSBIO-24-4105-T by Benjamin Lamp

“Dominance of Recombinant DWV Genomes with Changing Viral Landscapes as Revealed in Two National US Honey Bee and Varroa Mite Surveys”
Key Messages:

This study sequenced samples from the Bee Informed Partnership's bee health monitoring program. The viromes of commercially kept bees and their ectoparasitic mites were analyzed at the strain level using metagenomics. Viral sequences from 383 adult bees and 173 mites across 11 US states were identified, yielding 45 complete and 1,702 partial genomes. Notably, 312 Sinaivirus, 280 Sacbrood virus, and 135 DWV genomes were deposited in databases. Recombinant DWV genomes, comprising fragments from types A and B, were prevalent. The study raises concerns about the potential overestimation of type A prevalence due to current diagnostic assays, which may not distinguish between recombinant and type A genomes effectively. The nucleotide divergence in SBV and ABPV also causes false negatives in diagnostics. Overall, the study is methodologically sound, well-evaluated, and offers significant new insights. The finding that no major correlations exist between geographical origin and virus composition, likely due to the nationwide movement of bee colonies for pollination, is epidemiologically intriguing and supports the study’s conclusions. While the manuscript is well-written and comprehensive, addressing the following points would enhance its value:

Major Points for Consideration:

1. The use of a template-switching RT enzyme mix, as mentioned in the Methods (Line 302), may have led to an overestimation of recombinant genomes. This enzyme can add non-templated nucleotides to the 3'-end of cDNA, potentially generating artificial recombinants during hybridization with matching RNA genomes. The presence of both DWV-A and DWV-B variants in addition to recombinants should be verified, and this technical issue should be clearly acknowledged and discussed.

Response: The reviewer is correct to point out that we need to be certain that the method used do not create unwanted diversity such as recombinants. The template switching RT enzyme works through an addition of three cytosines when it reaches the 5’ end of the mRNA/viral genome molecule. Firstly, this enzyme does not readily fall off until it gets to the 5’ end. This is why it was used extensively in 5’ RACE experiments (SMARTer technology by Takara: https://www.takarabio.com/learning-centers/next-generation-sequencing/technology-and-application-overviews/smart-technology?srsltid=AfmBOoo0TYqpoeOwSZHOh_eTPK1zBZflVXiGA0oowu5ywEp_W281zkQa). Secondly, a template switching oligo that binds (hybridises) to the three cytosines is absolutely required (so it's targeted) and this is what is used for template switching. So it’s not a random switching event based on random genome hybridizations.

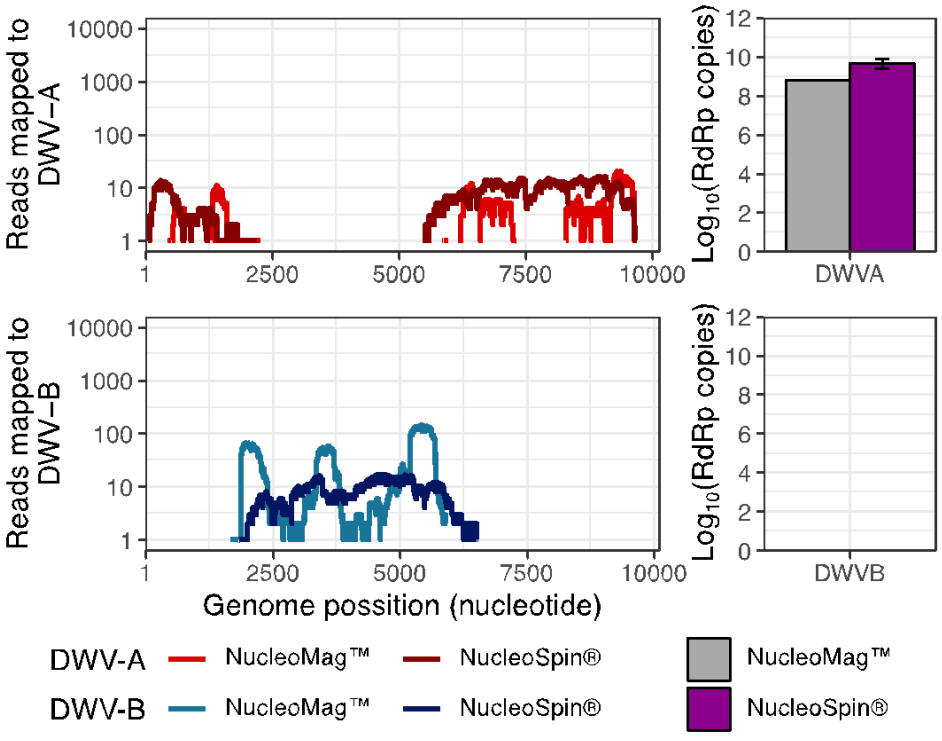
It must be remembered that template switching–based library preparation has emerged as a method of choice for single cell sequencing because of its relatively simple workflow and low RNA input requirements (Ziegenhain et al. 2017). It has also been adapted to sequencing small RNAs as well as fragmented RNAs from liquid

biopsies or microsomes (Rantalainen, 2018). Furthermore, template switching has been used in long-read single-molecule approaches with third-generation sequencing platforms, including isoform (Iso)-Seq (PacBio SMART-Seq) (Cartolano et al 2016) and nanopore RNA-Seq (Oxford Nanopore Technology) (Oikonomopoulos et al. 2016).

The reviewer should also be aware that we (Schroeder Lab) used this technology in other host-virus systems, e.g. sequenced PRRSV in pigs (Schroeder et al 2021) and MeV in humans (Yousaf et al., 2023). The Yousaf study used the Schroeder Lab's template switching sequencing pipeline to validate their Illumina based assemblies. NO recombinants were reported in this study.

Similarly, the LSV genomes that were recovered in the honey bees screen in Guyana did not have recombinants either (Schroeder et al 2022). And in our honey bee viromes associated with EFB (Hesketh-Best et al 2024), we did not recover many DWV recombinants, most were wild-type A or B (figure 2).

Finally, our entire sequencing workflow was further validated in the Nikulin et 2024 study. Here we demonstrated that RT-qPCR data concurs with ONT sequencing data generated where we confirmed the presence of recombinant genomes of DWV that were missed by qPCR detection (Figure S1 - see below for a copy). For this work we mapped reads against the same reference genomes for DWV-A and B genomes, as used in this study. We observed coverage of DWV-A that drops drastically before observing an increase in B, which then returns to a high coverage of DWV-A. We confirmed this region within the subsequent contigs generated from this sample and confirmed breakpoints (HyPhy-GARD), in a similar manner as performed within the current manuscript.



S1 Fig. Both extraction protocols are capable of identifying chimeric DWV genomes.
Colony # 8, Location: Hill, Year: 2021. Coverage histograms (left) represented as the number of reads mapped to two Deformed wing virus' strains from NCBI, (top) DWV-A (NC_004830.2) and (bottom) DWV-B/VDV-1 (NC_006494.1). Copy number of RNA-dependent RNA polymerase for the two DWV strains as detected by RT-qPCR (right).

Furthermore, for a random subset of the genomes generated in the current manuscript, due to the high labour intensity of this confirmation, we did observe multiple long-reads spanning the breakpoints identified within this present study - not presented within the manuscript due to brevity.

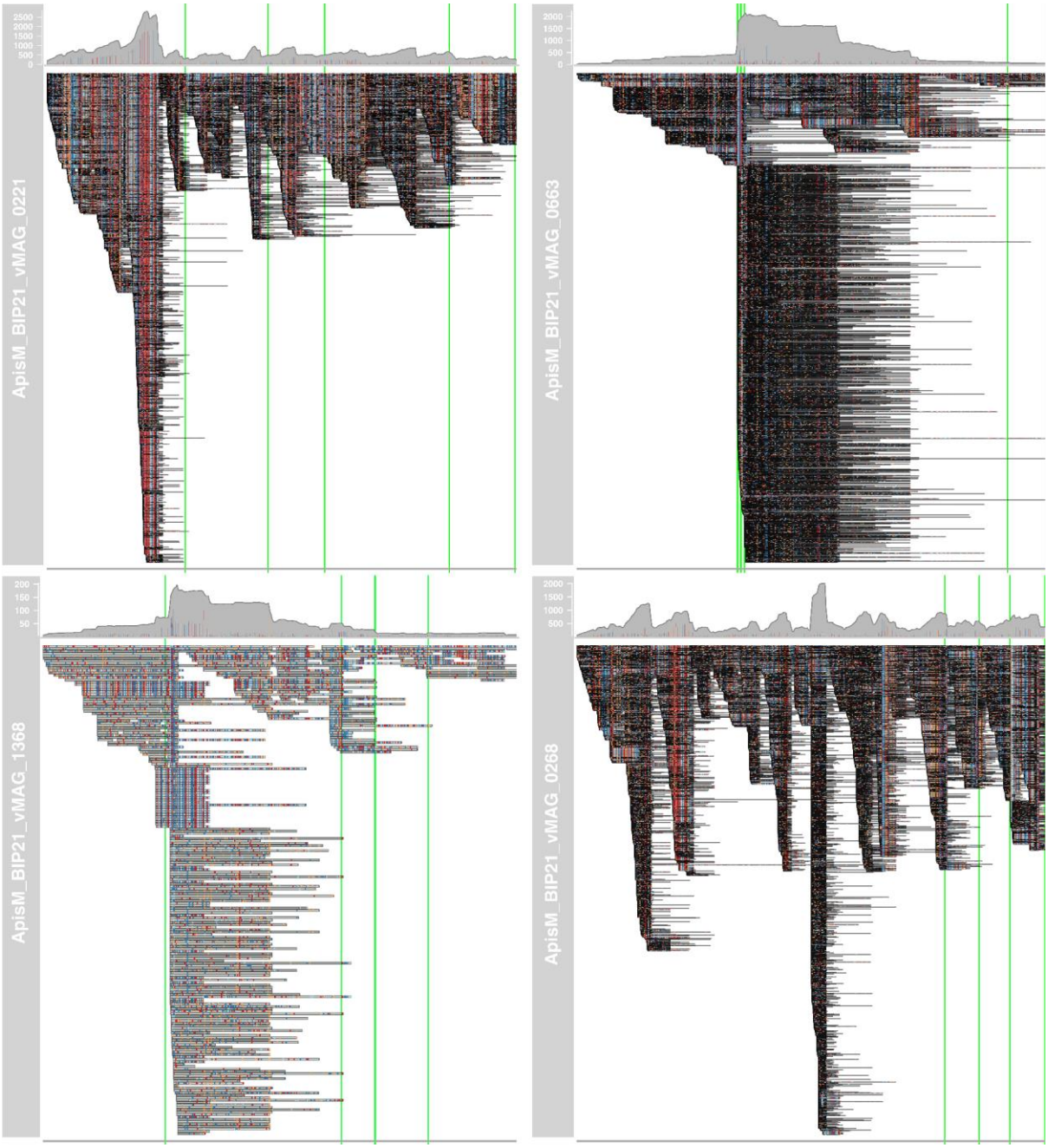


Figure. Validation of recombination breakpoint through the inspection of read mapping and coverage. We observe breakpoints as having multiple reads spanning 50 bp on either side of the breakpoint for recombinant genomes assembled in this study. Coverage (top) and read (bottom) plots of four recombinant genomes are presented here, with the HyPhy-GARD predicted breakpoint locations denoted as green lines.

Our recombinant hotspots are in agreement with prior literature (largely in the '5 UTR, Helicase (Dalmon et al 2017)). Discussed in the manuscript at lines 230-237. The arguments presented in this manuscript and here would suggest that while we can never rule out the possibility of chimeric assemblies, the genomic structures of our recombinant genomes are concurrent with the literature, our manuscript simply presents data that indicates recombinant genomes are more frequent than previously reported.

To conclude, we are confident that the method or technology used is not creating false recombinants.

Minor Points for Improvement:

1. Line 147: DWV has a monocistronic genome encoding a polyprotein in a single ORF. Thus, there are no 'intergenic regions' in the genome. Please clarify this point, possibly referring to "processing sites between mature proteins."

Response: We agree, thank you for the suggestion of this phrase, we have made the change throughout the manuscript.

2. Previous studies, including ours, have shown that despite their error-prone polymerase, DWV-A and DWV-B variants have remained stable over decades, with only minor mutations observed. This new study contributes to a vast number of full-genome sequences, expanding our understanding of DWV diversity. It would be valuable to present a multialignment of this primary dataset at the nucleotide level, alongside reference genomes, to create a master sequence for DWV-B and clarify the differences between these variants. Our group was able to present functionally tested genomes that were established in reverse genetic systems. As these are recent isolates, a comparison of the variants from the USA might be useful to draw epidemiological conclusions that include Europe. Of course, this does not mean that our studies must be cited!

Response: Given the scale of our data set (an example of a smaller data set shown above in major point 1), this type of figure would be impractical for a manuscript. Our sequences are available and such an alignment can easily be created by anyone. We also don't think that this has a practical application as viruses exist as quasi-species. So many of the nucleotide mutations can easily be lost or be part of a defective genome. For this reason we focused our attention on the protein sequence.

3. DWV-C is notably absent from the discussion. While our group has never detected this variant, it has been reported with high prevalence in the UK and USA. It is possible that our assays were not optimized for DWV-C detection. Was DWV-C detected using metagenomics

in this study? If not, why? Clarifying these findings and their implications would strengthen the manuscript.

Response: Correct, DWV-C was reported by us in the US in the Kevill et al 2019 paper. But this genome was not common and when it was found it was mainly in New Mexico. Moreover, type C was first reported in Devon, UK (also by the Schroeder Lab) in 2016 (Mordecai et al. 2016). Notably it was absent in the 2019 UK survey (Kevill et al 2019). So again it's clear that DWV-C is not abundant and not easy to find. It's also likely that honey bees are not its primary host. We found that it was dominant in small hive beetles (Brettell et al 2019). Interestingly, whole genome sequencing using Illumina in this study showed that the DWV-C in small hive beetles was a recombinant. Therefore, it's not unexpected that we did not find DWV-C in our study. Making this point could be a distraction as it is largely based on an hypothesis that we were not testing here.

That said, we have added the following statement to make readers aware of the lower limits of detection of our study, which includes both DWV-C and the recently reported A. mellifera solinvivirus-1 (AmSV1): *"Nor did we identify any DWV type C reads or contigs."* (line 226). We hope this emphasises to readers that not all viruses that have been detected in honey bees viromes were captured within this study, as this paragraph is concluded with the following statement: *"The absence of these uncommon viruses demonstrates a limit in detection of this study, possibly a result of our sequencing efforts not being expansive enough to detect low-titer viruses or our sampling occurred at a period of particularly low density."* (Lines 226-229)

Reviewer #3 (Remarks to the Author):

Data on virus sequence dynamics is important to understand their impact on hosts and the development of predictive models of virus-host interactions. Invertebrate RNA viruses are poorly characterized compared to those infecting human and other mammals. Therefore, results of this study, which investigated viromes in honey bees and Varroa mites, using nanopore high-throughput sequencing in several hundreds of colonies across the USA, are very valuable.

The major novel finding of this study is that recombinants between DWV-A and DWV-B are widespread in the USA. Novel types of recombinants (e.g. those with recombinant points in the helices gene) were detected. It should be emphasized, that DWV-B (VDV1) arrived to North America much later than DWV-A (see citation # 23, Ryabov et al 2017).

Response: Thank you for pointing this out. We modified the sentence in line 64 to emphasise this point: "Type B arrived much later in the US than type A[17], and until recently, type A was the most common variant in the US and potentially worldwide[12,23]" (citing Ryabov et a. 2017 buch earlier - now citation #17.

Therefore appearance of the DWV-A/DWV-B recombinants occurred in since early 2010s. This study made a "snapshot" of dynamic DWV complexes occurring in nature, providing a "real-life" example of virus evolutionary trends following arrival of an evolutionary distinct

variant. Considering recent arrival of DWV-B in the USA/Americas it would be important to mention the years of honeybee survey in the Abstract.

The study also convincingly showed differences of diversity of virus complexes in honey bees and Varroa mites which reflects differences between honey bee and Varroa mites in their ability to support replication of different pathogens (viruses).

Response: Thank you for this summary and positive feedback

I suggest to include a comparison of microbe compositions in honeybees and Varroa using samples derived from the same colonies and include result of this analysis as an additional panel in Fig. 1.

Response: Correct, we do have the microbial data BUT this is more complicated to present in the context of this study. This is because this is metatranscriptomics data and not the typical barcoding type 16S datasets. Moreover, the host genomes (nuclear and mitochondrial) also make up a significant proportion of the transcriptome. For this reason we chose not to include this analysis in this manuscript. This will be the subject of another paper.

Important point which is made in this MS is that commonly used (q)RT-PCR virus detection methods (which usually detect only a very small genome sections and/or do not produce sequencing data) do not allow to detect recombinant variants and may miss virus sequences due to mismatches between primers and viral RNA targets (L. 178-191). DWV-A in the USA was shown to have high variability (a paper reporting this could be cited https://doi.org/10.1371/journal.pbio.3000502). This study highlights the importance of wider use of unbiassed high-throughput sequencing methods to characterize virus populations.

Response: Agreed, we inserted the following sentence, "Furthermore, DWV-A has been observed to have high genetic variability, which may further complicate surveillance efforts[39]." (line 280), with citation.

I think that this MS contain sufficiently novel results to warrant publication in Communications Biology. I have only minor points which should be addressed:

Suggestions and comments:

Abstract. Indicate years of survey - DWV-B became widespread in the USA since mid 2010s.
Response: Done (Line 26)

L. 80 "we sequenced 71.8 million reads" -> "we generated 71.8 million sequencing reads"
Response: Done (Line 81)

L.83-88. Specify average (+- SD) length of viral reads used for the analysis / virus genome assembly.

Response: **The average read length of the sequence data has now been stated (mean= 415 bp +/- 124 bp) - Line 83.**

L. 147. Intergenic regions" - DWV has a single ORF encoding for a viral polyprotein which is cleaved to individual proteins. There are 5'- and 3'-terminal non-coding regions, but no "intergenic regions".

Response: **We have changed this to the region 'processing sites between mature proteins.' as per the suggestion of reviewer 2. Including in Table 3.**

Table 3.
Extend the table to include data on the sequence type for the 5' untranslated region and L-protein.

Response: **Assembly of a complete 5' UTR (and non contains complete 3' terminal ends of the genome) was uncommon. Table 3 reports the structure of only genomes with complete regions (excluding partial regions of the genomes). We have added the statement: "*The assembly of the 5' UTR was uncommon, and we did not obtain any DWV genomes with complete 3'-terminal non-coding regions.*" lines 151-152.**

To the table 3 legend we also added the following: "*We excluded the untranslated regions (UTR) of DWV from the following analysis and designated all genomic regions that are not mature protein sequences as 'Regions between mature proteins'*". See response to the comment below as to why L-proteins are not reported.

Indicate nucleotide positions of the sections "Vp1", Vp-2, Vp-3, Helicase, RdRp. What is the "intergenic region"? If it is a region between Halicase and RdRp, this section contains Vpg and Protease domains.

Response: **Due to the high number of genomes and to ensure replicability of these methods, we utilised InterProScan, which did not functionally classify the protease domains and L-proteins. Therefore we don't report the exact positional data.**

Fig. 3a.
Modify representation of DWV genomic RNA, indicate position of ORF, 5', 3' ends. Describe what Vp1 - 3, H., RdRp means. I suggest to modify schematic representation of DWV genome according to Ref. 84, (Lanzi et al 2006, Fig 3 A).

Response: **As stated above, these annotations are made utilising InterProScan, a tool that combines different protein signature recognition methods into a single output (see MM subsection '*Genome functional annotation and pSSN*'), and in the case of this study we utilised 6 distinct protein databases. We are disinclined to incorporate this suggestion as we do not feel a method based on the estimation of protein annotations using a single publication is a scalable or reproducible approach for metagenomic studies.**

This is indeed a limitation of our annotation approach, and as such we have added the following statement to make readers aware that not all known genomic genomes of

DWV are annotated: "*Our annotation method was limited by the fact that DWV L-proteins and proteases were not identified by InterProScan (#27, Lanzi et al 2006). Consequently, we are unable to comment on the frequency or occurrence of recombination within these genomic regions*" (lines 161-164).

L.145 / Table 3. "...thirteen distinct variants of recombinants [DWV] genomes ...". Actual nucleotide sequences of these 13 novel recombinants should be included as a Supplementary text to allow independent analysis.

Response: All full fasta file sequences of these contigs are present in the data repository, as stated in the Data and code availability statement. We feel that we included sufficient metadata to aid researchers to partition the recombinant genomes. Therefore we feel that an additional supplementary data file just for thirteen genomes is unnecessary in a manuscript already contains a large number of supplemental data files.