Review of Gavilán et al, submitted to Viruses 2023

Comments for authors

In their study, Gavilán et al investigate the molecular epidemiology of mumps in the Autonomous Community of Madrid (CM), primarily through the SH gene sequence analysis of 453 cases from 2015 to early 2020. A subset (90) was also subjected to sequencing of the non-coding region between the Matrix and fusion genes (MF-NCR) to explore its utility. This is an extension of their other recent works in which they looked at these same sequencing targets but on the national and international scale (references 18 and 19). In this study, the authors were particularly interested in seeing if these mumps sequencing tools could be used for transmission chain analysis at the local level. The sampling strategy for MF-NCR sequencing appears to have been those cases that were identified as possible clusters on the basis of commonalities in their SH gene variant (or haplotype), geographic and temporal distribution (Figure 2), in addition to cases identified as belonging to outbreaks by traditional epidemiology (Table 1). In general, the authors found that the additional sequence data obtained from the MF-NCR did not add additional phylogenetic resolution and in fact, the SH gene sequence provided more phylogenetic diversity (Figure 3), a finding that is of value to the community. Overall, I find that the study was well presented and have only minor suggestions for revision.

Minor comments:

1. Given that the conclusion of the study was that the MF-NCR sequence did improve the molecular resolution of similar mumps variants, I suggest that the title should be changed. Perhaps “Investigating” rather than “Revealing….with a combination of molecular tools.”
2. Likewise, the final sentence of the abstract is a conclusion that is based on the literature, not the current study. I think it should be moved into the introduction of the abstract, specifically line 27 immediately before the sentence starting with “The aim of this work…”.
3. The visual resolution of all figures needs to be improved. The text was illegibly small but was not made more legible by zooming in the figure. This includes the legends of all figures (variants) and the axes of Figure 1. In Figure 2, the triangles (derived haplotypes) could not be distinguished from the circles. In Figure 3, the text of the sequence names in the phylogenetic trees could not be clearly read. In all cases, this made it difficult to visually confirm the statements the authors made in the text regarding the findings.
4. In section 2.1, the authors describe the data within their database of RT-PCR confirmed mumps cases which included many demographic variables, none of which were used in this study, with the exception of geographic information. The description of the variables that are not used here should be removed as they are not within the stated aim of the current study. (As an aside, I encourage the authors to make use of this epi-lab linked data to conduct an in depth study of possible risk factors. In particular, linking the vaccination status to genotype or case status would be very interesting.)
5. Figure 1: Is the incidence provided the national incidence or only for CM? I suspect this is the national incidence. Please clarify in the figure caption.
6. Lines 174 – 180: it seems to me that the study authors are placing their findings within the broader national context which relies on data not presented in this study. Appropriate citations should be placed, and perhaps more appropriately, this should be moved to the Discussion.

7. In the geographic analysis (Figure 2), in addition to classifying SH gene sequences by variants and haplotypes, the authors use “derived haplotypes of SH variants.” Please provide the definition.
8. Please identify in the manuscript, and particularly in the caption for Figure 3, how long the two sequencing regions are (SH gene and MF-NCR).