

GENETIC DIVERSITY AND LOW-LEVEL INFECTION OF MALAWI POLYOMAVIRUS IN PEDIATRIC PATIENTS IN BRAZIL

Corresponding Author: Dr Svetoslav Naney Slavov

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 manuscript reports detection of Malawi polyomavirus, also known as polyomavirus 10, in nasopharyngeal samples from children living in the Sao Paulo region of Brazil. Of note, low viral loads were quantified in 18 samples, viral particles were detected in 4 samples by electron microscopy, and 6 whole genomes were reconstructed from Next-generation nucleotide sequence data.

General comments:

The study is well executed, and the data are clearly presented. The findings largely support what is already known about Malawi polyomavirus. The reporting of 6 new whole genomes is a modest contribution to the current state of knowledge. The identification of a recombination event in one genome is a novel finding, the significance of which is unclear.

Minor comments:

For orientation of the broader polyomavirus community, the authors should note that Malawi polyomavirus is also called polyomavirus 10.

Discussion, line 481-2: The authors state the 6 complete genomes they report are the largest number of full-length genomes reported to date. However, Rockett RJ et al reported 9 whole genomes in 2013 in Plos One (PMID: 23667518).

Figure 3: Please provide the nucleotide positions of regions 1, 2 and 3.

Figure 4 and 5: Is the reference bar for the electron micrographs labeled correctly as μm ? A bar in nm units would be more appropriate for this size virus. What is the diameter of the particles labeled as intact viral particles?

Reviewer #2

(Remarks to the Author)

The manuscript by Pereira et al. describes the genetic diversity of Malawi polyomavirus (MPyV) in young subjects from Brazil. This study extends the authors' previous work (J Med Virol. 2023 Apr;95(4):e28688. doi:10.1002/jmv.28688). While the work remains mainly descriptive regarding MPyV prevalence, it additionally identifies potential recombination events at the molecular level. Given the currently limited clinical significance of MPyV infection, I am not fully convinced about the overall impact and novelty of the manuscript.

Specific comments

Introduction

L69: Is MPyV truly an emerging virus, or is its apparent emergence mainly related to its relatively recent discovery?

L78: Please rephrase the sentence. It is awkward to start with "At the genome level" and continue with structural characteristics such as the absence of an envelope and icosahedral symmetry.

Materials and Methods

L195: Please clarify how the partial VP1 gene sequences were obtained.

Results

L303: Figures 2A and 2B could be moved to the supplementary material.

L307: Please clarify what is meant by “temporal analysis” in this context.

L358: What criteria were used to select the eight sequenced cases?

L398: What is the potential impact of the recombination event at the viral protein level? Additionally, what do the two “N” stretches in the MW57 sequence correspond to?

L399: “Figure X” should be corrected.

L429–438: I am not fully convinced by the identification of the observed particles. Immunogold labeling would help confirm the identity of these structures. The particles shown in Figure 1C resemble polyomavirus particles; however, in both images in panel B, the particles appear smaller than the expected ~45 nm diameter. Since the scale bars vary between images, it is difficult to accurately assess particle size. In addition, based on my experience with electron microscopy, which is a relatively low-sensitivity technique, there appears to be a discrepancy between the low viral loads detected by qPCR and the observation of numerous viral particles by EM.

Discussion

L507 and L524: Some repetition is present in these sections.

Reviewer #3

(Remarks to the Author)

719110 Review GENETIC DIVERSITY AND LOW-LEVEL INFECTION OF MALAWI POLYOMAVIRUS IN PEDIATRIC PATIENTS

This manuscript characterizes Malawi Polyomavirus (MWPvV) in pediatric patients in Brazil using whole genome sequencing (WGS), phylogenetic analysis of WGS and of VP1, and transmission electron microscopy (TEM).

General Comments [Ln is line number “n”]

1. Please provide more information in methods that to clarify quantitation.

a. PCR results are expressed as quantities per reaction (L143). Authors start with 2 ml of swab suspension. What was the volume used for extraction? What was the volume of the elution buffer? How much of this was added to the PCR tube?

b. TEM samples were prepared from 700 microliters of swab suspension that was concentrated into 20 microliters. L465 partially degraded particles were also observed.

c. Usually, one needs 10^6 to 10^7 particles per ml to see particles in TEM, yet PCR reactions indicated only 266 copies per reaction (high of 2.3×10^3). Please explain this discrepancy.

2. References throughout the text need better formatting. Some of the references are listed by number (example L76) while others are not (example L28). Some of the ones that are not don't appear in the reference list.

3. [L146-151; L307-311; L342-353] Respiratory co-infection analysis is weak. Only SARS-CoV2, Flu A and Flu B were assayed. Other common viruses like Rhinoviruses, RSV, Adenovirus, parainfluenza virus and HMPV were not assayed. Most of these viruses are cyclical. The authors need to relate peak months and off-season months of these other respiratory infections in Brazil relative to the dates of swab collections and to those swabs that were positive or negative for these other respiratory viruses.

4. The phylogenetic analysis of whole genome sequences and partial VP1 sequences and recombination analysis are a strong point of the manuscript.

5. Inclusion criteria are not clear enough. (L105-108; Figures 1B and 1C, and Supplement Figures 1 and 2; L269-278)

Inclusion criteria are not clear enough. Explain the differences in severity of presentations in the outpatient versus inpatient wards. What does “asymptomatic” mean and why would such children be in either ward? The sample size may have been sufficient, but it is not clear that the sample pool is representative.

6. L289-293 and Fig 1 A. Please indicate on the map with arrows the circles representing the patients with MWPvV57 and MWPvV105 and perhaps the other Genotype 1 isolates with a common symbol since they may represent a local transmission. It would be interesting to see to what extent the positive individuals and the Genotype 1-positived individuals were dispersed among all participants.

7. L415-422. Remember polyoma viruses are circular so that regions 1 and 3 are continuous. Perhaps, label the regions in Figure 3A since in your analysis they are not actually continuous since there are excluded sites between them in your analysis.

8. L459-469; L474 ; L539-540 Authors observed low titers in oropharyngeal tissue. Can you assume limited shedding. What is happening in blood, stools and urine which weren't measured? “respiratory origin” the oropharynx is part of both the upper respiratory tract and the upper gastrointestinal tract. In the example of poliovirus infections there is a very short transitory infection of the oropharyngeal cavity followed by viremia and then large-scale excretion if infection spreads to the gastrointestinal tract. Polyoma virus is activated in kidney transplant patients and titers are much higher than in blood. Were nasal swabs considered?

9. In Figure 3A, one of the excluded areas is the non-coding regulatory region (NCRR). This region undergoes major recombination in both BKV and JCV when replication is enhanced during activation or re-activation. Did you have any indication of any recombination in the NCRR of any of your isolates?

Specific comments [Ln is line number “n”]

719110 L2 Please add “in Brazil” to the end of the title.

L46; L481-483. Please explain “...the largest MWPvV genomic data set.” Authors provide WGS data for 6 of their isolates in figure 1D where there are an additional 24 whole genome sequences from other sources (see fig 1D and L174). One of these sources reported 9 whole genome sequences.

3. L109-111 "Oropharyngeal swabs were suspended in 2 mL of 1X sterile PBS and homogenized by vortexing prior to aliquoting for nucleic acid extraction." Suggest replacing "homogenized by vortexing" with "vortexed"
4. L209-212. State that the comparisons with BKV and JCV are only in the figures in the supplement as in L371.
5. L226 be more specific for "...multiple algorithms.". Which ones
6. L393-396; L496 correct "...distinct lineages..." in the next paragraph you clearly show that the isolate is a mosaic of two genotypes (2 and 3) with the majority of the nucleotides from Genotype 3, hence its position within genotype 3 on the figure.
7. Figures 4 and 5 and legends to figures. Please indicate which sub-figures (A to D) belong to which viral isolate, i.e., Figure 4A and 4B are MWPyV57, etc.
8. L450-451 "... in oropharyngeal swabs of Brazilian pediatric..."
9. L455-457 How does month of isolation relate to age-dependent increase in seroprevalence?
10. KL471 "MWPyV detection was predominantly observed in oropharyngeal cells of asymptomatic children "
11. L487-488 The genetic similarity to other Deltapolyomaviruses confirmed the membership of the isolates in this group
12. L522; L525-527 Why "ancient"? The recombination probably occurred before the specific individual was infected since there is no evidence of the rest of the genotype 2 sequence. However, there is no evidence when the recombination event occurred elsewhere.
13. L697 I was unable to see any red highlights.
14. L735 which "symbols"?

Reviewed by Prof. Lester M. Shulman

Open Access This Peer Review File is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

In cases where reviewers are anonymous, credit should be given to 'Anonymous Referee' and the source.

The images or other third party material in this Peer Review File are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

To view a copy of this license, visit <https://creativecommons.org/licenses/by/4.0/>

RESPONSE TO THE REVIEWERS COMMENTS/CRITICISMS

Reviewer #1 (Remarks to the Author):

The manuscript reports detection of Malawi polyomavirus, also known as polyomavirus 10, in nasopharyngeal samples from children living in the Sao Paolo region of Brazil. Of note, low viral loads were quantified in 18 samples, viral particles were detected in 4 samples by electron microscopy, and 6 whole genomes were reconstructed from Next-generation nucleotide sequence data.

General comments:

The study is well executed, and the data are clearly presented. The findings largely support what is already known about Malawi polyomavirus. The reporting of 6 new whole genomes is a modest contribution to the current state of knowledge. The identification of a recombination event in one genome is a novel finding, the significance of which is unclear.

We are grateful for the positive comment of the reviewer regarding our study.

We agree that the findings in our study are, in general, consistent with the current understanding of the biology of MW Polyomavirus, MWPvV (Polyomavirus 10). However, genomic information of MWPvV remains very limited especially from geographically underrepresented populations including Brazil. Additionally, the samples analyzed in our study, represent a rare and clinically relevant pediatric cohort and therefore do not reflect the general pediatric population investigated in community-based studies including the study by Rockett RJ et al., 2013.

We acknowledge that we committed an error stating that our study documents the largest number of MWPvV genomes. In fact, although the Rockett RJ et al., study reports a higher number of genomes (9 complete genomes), our work provides genomic data from distinct clinical settings and a specific population which was recruited in a public hospital in the State of São Paulo, Brazil. Beyond, reporting 6 complete genomes and identifying what appears to be the first recombination event described for MWPvV, our study also presents to our knowledge the first electron microscopy images of MWPvV directly obtained from clinical specimens (electron microscopy images of the expressed VP1 proteins in the form of MWPvV VLP are available). Altogether, our findings contribute complementary evidence regarding the genomic diversity, evolution, and morphological characterization of MWPvV during human infection.

We revised the “Discussion” section, highlighting the novelty and the significance of these aspects regarding MWPvV infection. The following paragraph was added:

Lines 608-618, Page 19: “... This study provides new genomic and morphological insights of MWPvV from a clinically relevant pediatric cohort in Brazil, a population that remains largely underrepresented in current polyomavirus research. In addition to generating six complete genomes from a distinct clinical setting, the performed genomic analysis identified what appears to be the first recombination event in MWPvV providing a better understanding of viral evolutionary dynamics. Furthermore, we provide ultrastructural evidence of MWPvV particles directly observed in clinical specimens. Collectively these data contribute complementary evidence regarding MWPvV genomic diversity, evolution and morphology during human infection, while highlighting the importance of broader genomic surveillance to refine phylogenetic relationships, identify novel lineages, and clarify the circumstances under which MWPvV may acquire clinical significance. ...”

Minor comments:

For orientation of the broader polyomavirus community, the authors should note that Malawi polyomavirus is also called polyomavirus 10.

We are grateful for the valuable recommendation of the reviewer. Accordingly, we have revised the manuscript to indicate at first mention that MWPyV is also known as Human polyomavirus 10 (HPyV10) and we have included this terminology where appropriate in the manuscript text.

Line 69, Page 3: "...also known as human polyomavirus 10..."

Discussion, line 481-2: The authors state the 6 complete genomes they report are the largest number of full-length genomes reported to date. However, Rockett RJ et al reported 9 whole genomes in 2013 in Plos One (PMID: 23667518).

We thank the reviewer for identifying this error. We agree that the statement was inaccurate. Rockett RJ et al., 2013, reported nine complete MWPyV genomes which exceeds the six genomes analyzed in our study. We have corrected the text accordingly. We would like to note however, that our study analyzed samples obtained only from pediatric patients attended in a public Brazilian hospital representing a distinct clinical population. Despite the lower number of genomes obtained compared to the above-mentioned study, the samples tested contained generally low viral loads and originated from a hospital-based setting. Therefore, the genomic data presented in our study presents important information regarding MWPyV diversity in a hospitalized pediatric population. We have revised our manuscript accordingly to accurately reflect these distinctions and avoid overstating the number of generated complete MWPyV genomes.

The following modifications were performed in the manuscript text:

We deleted the overstatement: which is the largest number of full-length genomes reported to date...

Lines 546-547, Page 17: we added the following explanation, instead of the "largest genome number dataset" "...These data expand the available genomic data for MWPyV from Brazil and support the circulation of genetically distinct viral lineages in the studied population. ..."

Figure 3: Please provide the nucleotide positions of regions 1, 2 and 3.

We thank the reviewer for the valuable suggestion. We agree that providing the nucleotide coordinates of genomic regions improves the clarity and interpretability of the genomic analysis. Accordingly, we have revised Figure 3 to include nucleotide positions of the identified regions in a linearized representation of the viral genome. Because polyomavirus genomes are circular, we have also included a circular representation highlighting the recombination breakpoints and corresponding recombinant genomic fragments, providing a more accurate visualization of the recombination events. For more information, please refer to the modified Figure 3.

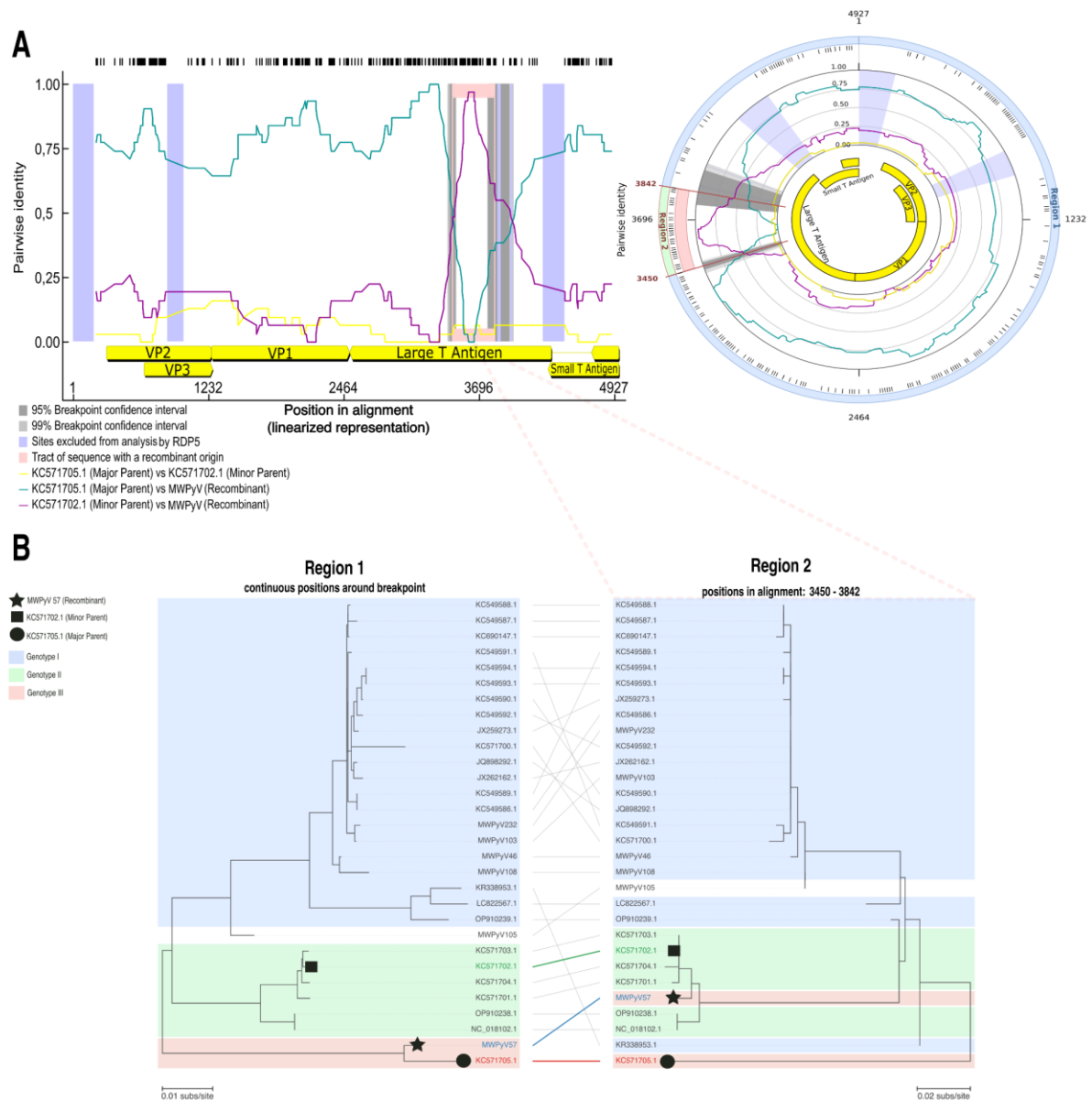


Figure legend:

Figure 3. Suggestive recombination event in the sample MWPvV57. **A.** Pairwise nucleotide identity plot generated in RDP5 showing similarity patterns between putative recombinant genome MWPvV57 and its inferred parenteral sequences across the complete MWPvV genome alignment. The y-axis indicates pairwise nucleotide identity, and the x-axis represents genomic position according to the alignment. Colored lines depict comparisons between MWPvV57 and the major parenteral sequence (KC571705.1) and the minor parenteral sequence (KC571702.1). Shaded vertical bands indicate the 95% and 99% breakpoint confidence intervals inferred by RDP5, whereas light-blue regions correspond to sites excluded from the analysis by RDP5. The predicted recombinant tract is highlighted, and the annotated coding regions are shown below the plot. Circular representation of the recombinant MWPvV genome showing region 1 which encompasses the continuous genomic sequence around the recombination breakpoint, and region 2, corresponding to the alignment positions 3450-3842. **B.** Maximum-likelihood phylogenetic trees reconstructed from two genomic partitions defined according to the inferred recombination breakpoint: Region 1 comprising continuous genomic sequence outside

the recombinant fragment, and Region 2, corresponding to the breakpoint-associated fragment (positions 3450-3842). The putative recombinant variant (star) shows incongruent phylogenetic placement between two regions, clustering with two different parenteral lineages depending on the genomic fragment analyzed. Genotype assignments (I-III) are indicated by background shading, and the symbols identify the inferred major and minor parenteral sequences.

Figure 4 and 5: Is the reference bar for the electron micrographs labeled correctly as μm ? A bar in nm units would be more appropriate for this size virus. What is the diameter of the particles labeled as intact viral particles?

We thank the reviewer for examining the electron micrographs. The reviewer is correct that the scale bars were incorrectly labelled as μm . The scale bars are in fact expressed in nm, and this has been corrected in the new figures 3 and 4 as well as in the corresponding figure legend. In addition, we measured the diameter of the particles identified as intact viral particles. We observed particles ranging from approximately 31.0 to 46.84 nm (mean/SD= 36,6 / 3,9 nm) which is consistent with the expected size of polyomaviruses. This information has now been added to the results section and to figure legends for clarity.

We added the following information into the results section:

Lines 452-455, Page 14: "... Measurements of viral particles revealed diameters ranging from 31.0 to 46.8 nm (mean $\pm$ SD, 36.6 $\pm$ 3.9), consistent with the expected size reported for polyomaviruses. However, particles smaller than the canonical $\sim$ 45 nm were also observed most likely representing partial or incomplete virions. ..."

Reviewer #2 (Remarks to the Author):

The manuscript by Pereira et al. describes the genetic diversity of Malawi polyomavirus (MPyV) in young subjects from Brazil. This study extends the authors' previous work (J Med Virol. 2023 Apr;95(4):e28688. doi:10.1002/jmv.28688). While the work remains mainly descriptive regarding MPyV prevalence, it additionally identifies potential recombination events at the molecular level. Given the currently limited clinical significance of MPyV infection, I am not fully convinced about the overall impact and novelty of the manuscript.

We are grateful to the reviewer for the comment. We agree that the clinical significance of MWPyV is still unknown and poorly explored and therefore the immediate clinical correlation of the presented findings is not possible at this stage. Given the well-established role of several human polyomaviruses in clinically significant diseases, particularly among immunocompromised individuals, the potential clinical impact of MWPyV warrants further investigation and may become clearer as additional epidemiologic, genomic and clinical data become available.

We would like to clarify that our previously performed study (J Med Virol. 2023 Apr;95(4):e28688. doi:10.1002/jmv.28688) was not designed to investigate MWPyV epidemiology or genetic diversity in specific patient groups. In that study, we detected incidentally through viral metagenomics two pediatric specimens positive for MWPyV DNA with elevated viral load during the SARS-CoV-2 Omicron wave, for which two complete genomes were recovered. This unexpected finding and the complete lack of information regarding the epidemiology and molecular data for MWPyV in Brazil, motivated the

present study, which was specifically designed to investigate the prevalence, genomic diversity, and molecular characteristics of MWPyV in a cohort of pediatric patients in a public Brazilian hospital. Beyond the expansion of the number of MWPyV genomes, the present study identified a putative recombination event and to our knowledge provides the first electron microscopy images of MWPyV directly observed in clinical samples. We acknowledge that although these findings did not establish causal relationship of MWPyV with the presence of clinical symptoms they contribute new information regarding the diversity, evolution and morphological characterization of this still poorly understood human polyomavirus. To address the reviewer's concern, we have revised the "Discussion" section to better clarify the scope, novelty and limitations of our findings:

Lines 608-618, Page 19. We added a paragraph in the "Discussion" section which emphasizes the novelty of the study: "... This study provides new genomic and morphological insights of MWPyV from a clinically relevant pediatric cohort in Brazil, a population that remains largely underrepresented in current polyomavirus research. In addition to generating six complete genomes from a distinct clinical setting, the performed genomic analysis identified what appears to be the first recombination event in MWPyV providing a better understanding of viral evolutionary dynamics. Furthermore, we provide ultrastructural evidence of MWPyV particles directly observed in clinical specimens. Collectively these data contribute complementary evidence regarding MWPyV genomic diversity, evolution and morphology during human infection, while highlighting the importance of broader genomic surveillance to refine phylogenetic relationships, identify novel lineages, and clarify the circumstances under which MWPyV may acquire clinical significance. ..."

Lines 534-538; 539-540, Page 16,17: We added a phrase pointing out the limitation to link the detection of MWPyV DNA with respiratory disease: "...In this context, the detection of MWPyV in respiratory samples alone does not establish a causal association with respiratory disease. The limited number of symptomatic MWPyV-positive children and the concurrent co-circulation of multiple respiratory viruses during the study period preclude definitive conclusions regarding its pathogenic role. Collectively, our findings are consistent with the hypothesis that MWPyV is a part of the human virome, although its biological and clinical significance requires further investigation. ..."

Specific comments

Introduction

L69: Is MPyV truly an emerging virus, or is its apparent emergence mainly related to its relatively recent discovery?

We thank the reviewer for this important observation. We agree that the term "emerging virus" may not be the most appropriate term to describe MWPyV. Like other human polyomaviruses, MWPyV is probably an ancient virus that has co-evolved with its host over an extended evolutionary timescale. Its apparent emergence is more likely related to its relatively recent discovery through advancements in next-generation sequencing rather than to a recent introduction event in the human population.

Indeed, a similar situation has been observed for many other human polyomaviruses that have been identified over the last two decades. While BKPyV and JCPyV were for many years the only recognized human polyomaviruses, advances in molecular detection techniques and metagenomic studies led to the discovery of additional polyomavirus species, including KI polyomavirus, WU polyomavirus, Merkel cell polyomavirus, Trychodysplasia spinulosa polyomavirus, human polyomavirus 6, human

polyomavirus 7, human polyomavirus 9, Saint Louis polyomavirus, New Jersey polyomavirus, and Lyon IARC polyomavirus. Therefore, MWPyV should be regarded as a recently identified and poorly characterized human polyomavirus rather than a truly emerging virus.

To avoid potential confusion, we have revised the “Introduction” section and replaced the term “emerging virus” with “recently identified human polyomavirus” (or similar wording where appropriate).

Lines 69-70, Page 3: “...a recently identified virus...”

L78: Please rephrase the sentence. It is awkward to start with “At the genome level” and continue with structural characteristics such as the absence of an envelope and icosahedral symmetry.

We thank the reviewer for the valuable observation. We agree that the original wording was imprecise because it combined genomic and structural characteristics in the same phrase. The sentence has been revised for clarity in the following manner:

Lines 78-79 Page 3: “MWPyV is a non-enveloped virus with an icosahedral capsid and a ~ 5 kb circular double-stranded DNA genome.”.

Materials and Methods

L195: Please clarify how the partial VP1 gene sequences were obtained.

The reference coding sequence of the VP1 protein was retrieved from the nucleotide database of the NCBI. Using Biopython (v1.76), the corresponding GenBank flat files of the complete sequences (24) were downloaded and parsed to identify CDS feature annotations, allowing the extraction of the nucleotide sequence associated with the VP1 gene. Additionally, we included 7 complete VP1 coding sequences retrieved from the GenBank. To enable local homology searches, a BLAST nucleotide database was created from the multifasta file containing all consensus genomes using the NCBI BLAST+ makeblastdb command. The reference VP1 nucleotide sequence was then used as a query in a local BLASTn search against the consensus genome database.

For each consensus genome (sseqid), only the alignment with the highest bit score was retained, ensuring the selection of the most likely VP1 homologous region. The genomic coordinates obtained from the selected alignments were subsequently used to extract the corresponding VP1 regions from each consensus genome using a custom Python script. The resulting VP1 sequences were exported in FASTA format for downstream analyses.

We included both Python scripts (extract the VP1 CDS from the gbff file and extract VP1 CDS from consensus sequences) in the GitHub repository. We also added the following explanation in the “Materials and Methods” section:

Lines 204-209, Page 7: “... For the VP1 phylogenetic analysis, coding sequences were identified by querying a reference VP1 sequence obtained from NCBI against a local BLAST database built from the consensus genomes. For each genome, the best-scoring BLASTn hit was selected, and the corresponding VP1 region was extracted using custom Biopython v1.76. Extracted sequences were exported in FASTA format for downstream analyses. All scripts are available in the GitHub repository (https://github.com/gmcampos27/MWPyV_Analyses)....”

Results

L303: Figures 2A and 2B could be moved to the supplementary material.

We moved these figures to the supplementary data, named as Supplementary figure 3. Figure 2C was included as panel 1B in Figure 1.

L307: Please clarify what is meant by “temporal analysis” in this context.

We thank the reviewer for this comment. We agree that the term “temporal analysis” may be too broad and may imply more detailed time-series or seasonal analysis. Our intention was to simply describe the monthly distribution of MWPv DNA positive samples throughout the study period. To improve the clarity, we revised the text accordingly:

Lines 331-333, Page 10: “... Analysis of the monthly distribution of the MWPv positive cases revealed detections throughout 2024, with cases identified across multiple non-consecutive months and no clear seasonal pattern. ...”

L358: What criteria were used to select the eight sequenced cases?

We thank the reviewer for this question. The eight samples selected for sequencing were chosen based on their TaqMan® real-time PCR Ct values and viral load, as viral load is a major determinant for successful whole genome recovery. Two samples exhibited relatively low Ct values (Ct values ~ 29 and 30), while the other selected samples showed Ct values up to 33. The additional samples were included because we considered the sequencing depth achievable by the Illumina ISeq platform which could still allow recovery of complete or nearly complete genomes despite the low viral loads.

Samples with Ct values > 33 were not selected because the viral loads were considered unlikely to yield sufficient genomic coverage for reliable assembly of the complete MWPv genome. We have clarified the selection criteria in the “Materials and Methods” section:

Lines 165-168, Page 5: “... Samples selected for sequencing were prioritized based on the viral load as estimated by the cycle threshold values (Ct). All samples with Ct < 33 were included for sequencing and samples with higher Ct values were not considered due to the expected low probability of recovering complete viral genomes....”

L398: What is the potential impact of the recombination event at the viral protein level?

We thank the reviewer for raising this important point. To assess the potential impact of the recombination event at the viral protein level, we performed additional analyses at both sequence and structural levels.

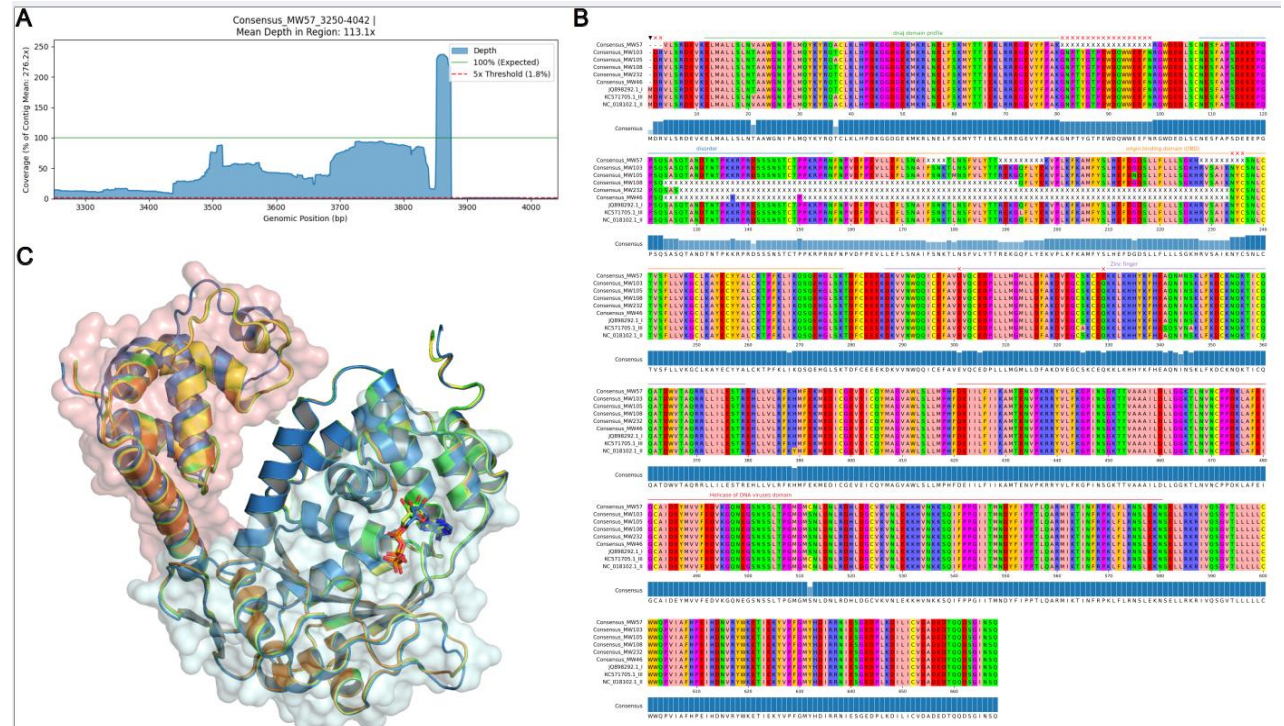
First, we evaluated the sequencing support for the recombinant region by examining read depth and genome coverage using mapped reads. The coverage plot showed consistent read support across the inferred recombinant breakpoint region, with a minimum depth of approximately 5×, supporting that the identified recombinant sequence was not an artefact of genome assembly or sequencing.

Next, we extracted the coding sequence corresponding to the recombinant genome and compared its translated amino acid sequence with those of MWPvV genotypes 1, 2 and 3 available in the NCBI database. Overall, the amino acid sequence of the recombinant strain was highly conserved relative to the reference genotypes. Importantly, no major amino acid changes, premature stop codons or disruptions of the predicted Large T antigen open reading frame were observed in the recombinant region.

Because the inferred recombination event involved the genomic region encoding the Large T antigen, we further investigated whether it could affect the structural integrity of functionally relevant domains. Three-dimensional structural predictions of the zinc-finger and helicase domains were generated using AlphaFold 3 for the recombinant sequence and representative sequences from genotypes 1, 2 and 3. The predicted structures were highly similar across all analyzed sequences, indicating preservation of the overall folding and structural organization of these domains.

Taken together, these findings suggest that, although the recombinant genome is supported by sequencing reads and points out a probable recombination event, it does not appear to cause major changes in the amino acid composition or predicted three-dimensional structure of the zinc-finger and helicase domains of the Large T antigen. Therefore, this possible recombination event did not affect the structural integrity of these protein regions. We have added a supplementary figure and we have discussed these findings in the revised version of the manuscript:

We added the following supplementary figure in the section “Supplementary figures” (Supplementary figure 7):



Supplementary Figure 7. Sequence and structural characterization of the recombinant Large T antigen from the MWPvV57 isolate. (A) Normalized sequencing depth across the inferred recombinant region (positions 3450-3842) of the MWPvV57 genome. Sequencing reads were mapped

to the assembled viral genome, and per-base coverage was calculated using SAMtools and the *pysam* Python library. Coverage values were normalized to the expected mean coverage of the corresponding contig, estimated from the number of mapped reads, average read length, and contig length. The green horizontal line indicates the expected normalized coverage (100%), whereas the red dashed line corresponds to the equivalent of 5× sequencing depth. Continuous read support across the inferred recombination breakpoint supports the authenticity of the recombinant sequence and argues against an assembly artefact. **(B)** Multiple amino acid sequence alignment of the Large T antigen from the recombinant MWPvV57 isolate and representative Malawi polyomavirus genotypes I (JQ898292.1), II (NC_018102.1), and III (KC571705.1). Coding sequences (CDSs) were translated and aligned using Clustal Omega. The recombinant region comprises exclusively synonymous nucleotide substitutions, resulting in complete conservation of the amino acid sequence, including the origin-binding, Zinc Finger, and Helicase domains. **(C)** Structural comparison of the Zinc Finger and Helicase domains predicted using AlphaFold 3. Surface representation highlights the Zinc Finger domain in **salmon** and the Helicase domain in **pale cyan**. Cartoon representations are colored according to genotype: MW57 (green), genotype I (JQ898292.1; sky blue), genotype II (NC_018102.1; yellow), and genotype III (KC571705.1; orange). Structural superposition reveals highly conserved three-dimensional conformations among all sequences, indicating that the recombination event does not alter the predicted structural organization of these functional domains.

We have added the following section in the “Materials and Methods” section:

Lines 246-266, Page 8: “... To evaluate whether the inferred recombination event could affect the encoded protein, complementary analyses were performed at the sequencing, sequence, and structural levels. Sequencing reads were mapped to the assembled viral genome (MWPvV57), and read support for the recombinant region was assessed by calculating per-base sequencing depth using SAMtools¹⁸ and the *pysam* Python library. To account for differences in sequencing depth among contigs, coverage was normalized to the expected mean contig coverage, estimated from the number of mapped reads, average read length, and contig length obtained from samtools idxstats. Normalized coverage profiles were generated using Matplotlib to verify continuous sequencing support across the inferred recombination breakpoint. The complete coding sequence of the LTA_g was extracted from the recombinant genome and from representative MWPvV genotype 1 (JQ898292.1), genotype 2 (NC_018102.1), and genotype 3 (KC571705.1) genomes available in the NCBI database. The nucleotide sequences were translated into amino acid sequences and aligned using Clustal Omega²⁹ to identify potential amino acid substitutions, insertions, deletions, premature stop codons, or other sequence alterations associated with the inferred recombination event. To investigate whether the observed sequence differences could affect protein structure, three-dimensional structural models were generated for the zinc-finger and helicase domains of the LTA_g. The amino acid sequences corresponding to each domain were extracted from the recombinant strain and from representative genotype 1, 2, and 3 sequences. Structural predictions were generated using AlphaFold 3³⁰, and the resulting models were visually compared to evaluate the conservation of the overall folding and structural organization of both functional domains. ...”

We have revised the discussion accordingly:

Lines 593-606, Page 18: “... An important question is whether this recombination event affects the structural integrity of MWPvV LTA_g. Notably, the recombinant segment contained exclusively synonymous nucleotide substitutions, resulting in no changes to the predicted LTA_g amino acid

sequence. Structural models generated using AlphaFold 3 for the zinc-finger and helicase domains showed highly similar three-dimensional conformations between the recombinant genome and representative parental genomes, consistent with the absence of predicted structural alterations in these functionally important regions (Supplementary figure 7). Together, these findings suggest that the recombination event did not affect the primary amino acid sequence or the predicted structure of LTA_g. This observation is consistent with evolutionary constraints acting on this essential viral replication protein. If the recombinant genome provides an evolutionary advantage, such an effect may be related to nucleotide-level features, including codon usage, mRNA stability, RNA secondary structure, or translation efficiency, rather than to changes in LTA_g protein structure. However, functional studies will be required to determine whether these synonymous substitutions have any biological consequences. ...”

Additionally, what do the two “N” stretches in the MW57 sequence correspond to?

We thank the reviewer for the careful observation of the complete genome sequence. The two stretches of “N” in the MWPyV57 genome correspond to regions for which no reliable nucleotide sequence could be obtained due to a lack of sequence coverage. The unresolved positions were therefore presented as “N” in the consensus sequence.

L399: “Figure X” should be corrected.

We corrected Figure XXX. We apologize for this error.

L429–438: I am not fully convinced by the identification of the observed particles. Immunogold labeling would help confirm the identity of these structures. The particles shown in Figure 1C resemble polyomavirus particles; however, in both images in panel B, the particles appear smaller than the expected ~45 nm diameter. Since the scale bars vary between images, it is difficult to accurately assess particle size. In addition, based on my experience with electron microscopy, which is a relatively low-sensitivity technique, there appears to be a discrepancy between the low viral loads detected by qPCR and the observation of numerous viral particles by EM.

We appreciate the reviewer’s careful assessment of the obtained ultrastructural data.

Based on the measurements across multiple micrographs, we observed that the viral particles show size between 31.0 and 46.8 nm (SD 36.6 + 3.9) using the software ImageJ (v. 1.54p, National Institutes of Health, Bethesda, MD, USA) which is consistent with the expected size of the polyomaviruses. The scale bars in the original figures were incorrectly labelled and have been corrected to nm.

As observed by the estimated particle size, our images also show particles smaller than the canonical ~45 nm. We likely contend that these smaller structures most likely represent partial or incomplete viral particles. It is common in ultrastructural studies of viral morphogenesis that a heterogeneous population of particles, including assembly intermediates and defective capsids, can occur. We updated the manuscript to discuss that these size variations are indicative of partial virions.

The point regarding discrepancy between the low viral load detected by qPCR and the frequency of particles observed via transmission electron microscopy (TEM), in our view, can be explained by two main factors: (1) since qPCR quantifies viral genomic DNA and TEM visualizes physical proteinaceous structures, an abundance of empty capsids or defective interfering particles would result in a high

particle count under the microscope despite the relatively low concentration of detectable viral DNA; (2) polyomaviruses often replicate in concentrated “viral factories” within the infected cells. While overall loads in swabs (in our case) or bulk tissue samples quantified by qPCR might be low, the probability of capturing these high-density focal points during TEM sampling can lead to observation of multiple particles in specific TEM sections.

We completely agree that the immunogold labelling is the gold standard for definitive MWPvV particle identification. However, after thorough search in the different antibody providers and to the best of our knowledge there are no commercially validated antibodies against MWPvV that are currently available for immunogold labelling. However, we believe that morphological characteristics, combined with biological context and our precise measurements, provide a reasonable foundation for the preliminary identification of the viral particles as polyomaviruses.

We revised the figures, legends and the “Discussion” section to clarify these points and to avoid overinterpretation of these findings. The following information have been added to the manuscript:

We added the performed measurements in the “Results” section:

Lines 452-455, Page 14: “... Measurements of viral particles revealed diameters ranging from 31.0 to 46.8 nm (mean $\pm$ SD, 36.6 ± 3.9), consistent with the expected size reported for polyomaviruses. However, particles smaller than the canonical ~ 45 nm were also observed most likely representing partial or incomplete virions. ...”

We added the following interpretation in the “Discussion” section:

Lines 497-511, Pages 15,16: “... Electron microscopy revealed morphologically complete polyomavirus-like particles in oral swabs, providing additional evidence for the presence of MWPvV in the oral cavity which can be regarded as a relevant size for viral presence in addition to the intestinal tract supporting oral-fecal transmission.⁷ The observed particles measured 31.0-46.8 nm in diameter, consistent with the expected size range of polyomaviruses. Although some particles appeared smaller than the canonical polyomavirus size and may represent incomplete or partially assembled virions, the combination of particle morphology, size measurements, and molecular detection supports their interpretation as MWPvV-associated viral particles. The apparent discrepancy between low viral loads detected by qPCR and particle frequency observed by transmission electron microscopy may be explained by the presence of empty or defective capsids, which can be visualized despite containing little or no viral genomic DNA, as well as by the focal accumulation of viral particles within localized replication sites. Although immunogold labelling would provide definitive confirmation of MWPvV identity, to the best of our knowledge no MWPvV-specific antibodies suitable for immunogold electron microscopy are currently available....”

Discussion

L507 and L524: Some repetition is present in these sections.

We are grateful for the careful observation of the reviewer. We agree that these are similar statements. For that reason, the statement in Line 524 was deleted and only the statement of Line 507 was maintained in the revised version of the manuscript.

We left the following phrase: **Lines 568-571, Page 17:** "... Recombination in polyomaviruses may be facilitated by cis-acting regulatory elements and bidirectional transcription across the viral regulatory region, which can promote homologous pairing and strand exchange independently of viral DNA replication³⁶. ..."

Reviewer #3 (Remarks to the Author):

719110 Review GENETIC DIVERSITY AND LOW-LEVEL INFECTION OF MALAWI POLYOMAVIRUS IN PEDIATRIC PATIENTS

This manuscript characterizes Malawi Polyomavirus (MWPvV) in pediatric patients in Brazil using whole genome sequencing (WGS), phylogenetic analysis of WGS and of VP1, and transmission electron microscopy (TEM).

General Comments [Ln is line number "n"]

1. Please provide more information in methods that to clarify quantitation.

a. PCR results are expressed as quantities per reaction (L143). Authors start with 2 ml of swab suspension. What was the volume used for extraction? What was the volume of the elution buffer? How much of this was added to the PCR tube?

We are grateful for the valuable comment of the reviewer. The extraction of viral nucleic acids was performed using the QIAamp Viral RNA Mini Kit (QIAGEN) following the manufacturers instruction. The sample volume used for extraction was 140 µL and the extracted nucleic acids were eluted in 60 µL nuclease-free water. For the qPCR assay, 6 µL of the extracted nucleic acids were used as PCR template in each reaction. We added this information in the "Materials and Methods" section:

Lines 129-131, Page 4: "... Viral nucleic acids were manually extracted from 140 µL of oral swab specimen using the QIAamp Viral RNA Mini Kit (QIAGEN) according to the manufacturer's instructions and eluted in a final volume of 60 µL. ..."

Lines 133-135, Pages 4,5: "... Reactions were performed in a final volume of 20 µL, containing 6 µL of extracted nucleic acids, 400 nM of each primer, and 200 nM of the hydrolysis probe, using the GoTaq® Probe 1-Step RT-qPCR System (Promega). ..."

b. TEM samples were prepared from 700 microliters of swab suspension that was concentrated into 20 microliters. L465 partially degraded particles were also observed.

We are grateful for the valuable observation of the reviewer. We agree that the partial viral particles observed may have resulted from the concentration process. This was changed in the revised version of the manuscript in the following manner:

Lines 448-449, Page 14: "... Viral particles were detected in all examined samples, although some appeared partially degraded. ..."

c. Usually, one needs 10e6 to 10e7 particles per ml to see particles in TEM, yet PCR reactions indicated only 266 copies per reaction (high of 2.3 x10e3). Please explain this discrepancy.

We thank the reviewer for this important observation. We agree that the transmission electron microscopy (TEM) generally requires high particle concentrations for reliable visualization. In first place, we would like to clarify that we chose the samples that showed higher viral loads, MWPyV57, 103, 105, 139 and 232 respectively. The viral loads were 179.5, 2346, 1037, 182.5 and 285 copies/reaction respectively. In this respect, we completely agree that samples with lower viral load were also included, and the explanation for the observed viral particles using TEM is given below.

The concern regarding the visualization of viral particles in samples with low viral load was also expressed by reviewer 2. This discrepancy in our opinion, can be explained by two main factors: (1) since qPCR quantifies viral genomic DNA and TEM visualizes physical proteinaceous structures, an abundance of empty capsids or defective interfering particles would result in a high particle count under the microscope despite the relatively low concentration of detectable viral DNA; (2) polyomaviruses often replicate in concentrated “viral factories” within the infected cells (aggregation of viral particles was also observed in our TEM analysis). While overall loads in swabs (as in our case) or bulk tissue samples quantified by qPCR might be low, the probability of capturing these high-density focal points during TEM sampling can lead to observation of multiple particles in specific TEM sections.

Additionally, prior to electron microscopy, all positive samples underwent multiple concentration steps using Amicon ultrafiltration devices, substantially increasing the concentration of viral particles for the final preparation. However, we acknowledge that the sole use of TEM does not directly confirms MWPyV particles and due to this we have added the following paragraph in the “Discussion” section:

Lines 497-511, Pages 15-16: “...Electron microscopy revealed morphologically complete polyomavirus-like particles in oral swabs, providing additional evidence for the presence of MWPyV in the oral cavity which can be regarded as a relevant size for viral presence in addition to the intestinal tract supporting oral-fecal transmission.⁷ The observed particles measured 31.0-46.8 nm in diameter, consistent with the expected size range of polyomaviruses. Although some particles appeared smaller than the canonical polyomavirus size and may represent incomplete or partially assembled virions, the combination of particle morphology, size measurements, and molecular detection supports their interpretation as MWPyV-associated viral particles. The apparent discrepancy between low viral loads detected by qPCR and particle frequency observed by transmission electron microscopy may be explained by the presence of empty or defective capsids, which can be visualized despite containing little or no viral genomic DNA, as well as by the focal accumulation of viral particles within localized replication sites. Although immunogold labelling would provide definitive confirmation of MWPyV identity, to the best of our knowledge no MWPyV-specific antibodies suitable for immunogold electron microscopy are currently available. ...”

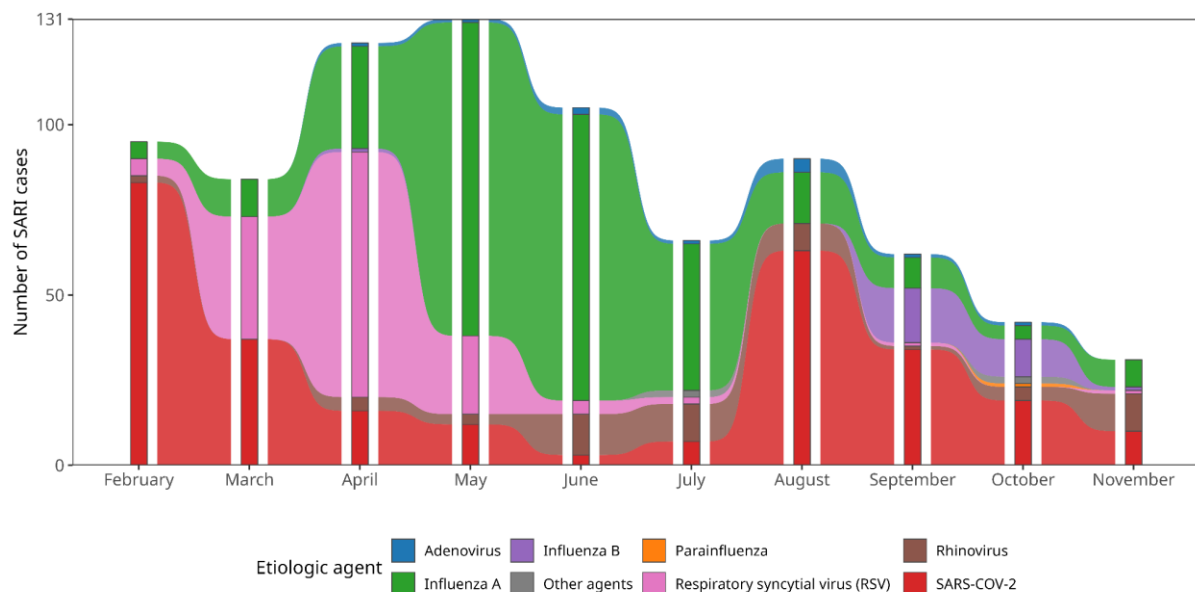
2. References throughout the text need better formatting. Some of the references are listed by number (example L76) while others are not (example L28). Some of the ones that are not don't appear in the reference list.

We thank the reviewer for pointing this out. We have carefully revised the reference formatting through the manuscript to ensure consistency. All the references are now presented using the journal-required format with numbers as superscripts.

3. [L146-151; L307-311; L342-353] Respiratory co-infection analysis is weak. Only SARS-CoV2, Flu A and Flu B were assayed. Other common viruses like Rhinoviruses, RSV, Adenovirus, parainfluenza virus and HMPV were not assayed. Most of these viruses are cyclical. The authors need to relate peak months and off-season months of these other respiratory infections in Brazil relative to the dates of swab collections and to those swabs that were positive or negative for these other respiratory viruses.

We thank the reviewer for this important comment. The respiratory virus screening in the positive for MWPvV children was limited only to SARS-CoV-2, Influenza A and Influenza B because these pathogens are primary targets of the respiratory surveillance program conducted among the Butantan Institute in collaboration with the São Paulo municipal health authorities.

Furthermore, epidemiological surveillance data from Ribeirão Preto city, showed that these viruses are the predominant respiratory pathogens circulating in the studied period, accounting for major seasonal peaks of respiratory infections (See figure below). Nevertheless, analysis of the publicly available data showed that other viruses including rhinoviruses, adenoviruses, respiratory syncytial virus, metapneumovirus and parainfluenza viruses also circulated but were not tested among the MWPvV positive patients. Importantly, only small proportion of the MWPvV positive children presented respiratory symptoms at the time of sample collection, although two of them showed positive results for SARS-CoV-2 RNA with high Ct values. Therefore, although the presence of additional respiratory co-infections could not be excluded because comprehensive respiratory panel testing was not performed, the likelihood that an undetected respiratory pathogen was responsible for a clinically apparent respiratory disease in these individuals remains limited. To address this reviewer concern, we have added a supplementary figure and complementation in the discussion section providing the temporal circulation patterns of major respiratory viruses during the study period in the city of Ribeirão Preto, where the study was conducted, based on publicly available epidemiological surveillance data.



Supplementary Figure 6. Temporal distribution of severe acute respiratory infection (SARI) cases by etiological agents in the city of Ribeirão Preto, the monthly distribution of laboratory-confirmed respiratory pathogens among SARI cases from February to November 2024. Flow widths are proportional to the number of cases attributed to each virus per month. The visualization highlights temporal fluctuations and the changing relative contribution of each etiological agent over time,

including SARS-CoV-2, influenza A and B viruses, respiratory syncytial virus (RSV), rhinovirus, adenovirus, parainfluenza viruses, and other respiratory pathogens.

We have also added the following text in the “Discussion” section:

Lines 521-531, Page 16: “... A limitation of our study was that the MWPyV-positive samples were tested only for SARS-CoV-2, Influenza A and B which were predominant respiratory viruses circulating the city of Ribeirão Preto, according to municipal epidemiological surveillance data. Although these viruses accounted for major circulation peaks observed during the sampling period, other respiratory pathogens, including principally respiratory syncytial virus but also adenoviruses, rhinoviruses, parainfluenza viruses, and human metapneumovirus were also observed albeit at lower frequencies. The temporal dynamics of these respiratory viruses is presented in Supplementary Figure 6. Therefore, although the presence of additional respiratory infections cannot be excluded because a comprehensive respiratory panel was not performed, the possibility of viral co-infections should be considered when interpreting the clinical significance of MWPyV. ...”

4. The phylogenetic analysis of whole genome sequences and partial VP1 sequences and recombination analysis are a strong point of the manuscript.

We thank the reviewer for this positive assessment. We are pleased that the reviewer recognizes the phylogenetic analysis of the whole genome and partial VP1 sequences, as well as the recombination analysis as strengths of our study. These analyses provided important insights into the genetic diversity and evolutionary dynamics of MWPyV circulating in our cohort. In particular, the recombination analysis contributed to the identification of putative recombination events, which to our knowledge represent the first evidence of recombination reported to MWPyV.

5. Inclusion criteria are not clear enough. (L105-108; Figures 1B and 1C, and Supplement Figures 1 and 2; l269-278) Inclusion criteria are not clear enough. Explain the differences in severity of presentations in the outpatient versus inpatient wards. What does “asymptomatic” mean and why would such children be in either ward? The sample size may have been sufficient, but it is not clear that the sample pool is representative.

We thank the reviewer for this important and insightful observation. We completely agree that the inclusion criteria, the clinical categorization of the participants, and the limitations regarding sample representativeness required further clarification.

To address these points, we have thoroughly revised the "Materials and Methods" section to provide a more detailed description of patient recruitment and definitions:

1. Patient recruitment: Children were enrolled from both inpatient and outpatient services in the Pediatric Department of HC-FMRP-USP irrespective of the presence of respiratory symptoms. Outpatients primarily consisted of children attending scheduled routine consultations and specialized clinical follow-up visits. Inpatients were hospitalized for a variety of acute medical or surgical conditions completely unrelated to the objectives of this study.

2. Definition of "Asymptomatic": The classification of clinical presentation and underlying disease was performed independently. Patients were first categorized according to the presence or absence of

clinical symptoms at the time of evaluation. The term “asymptomatic” refers to children who did not present any of the clinical symptoms included in the study “classification” (e.g. respiratory, gastrointestinal, neurological, or other recorded symptoms) during clinical assessment. Independently, patients were also categorized according to previously established medical diagnoses (16 disease categories) or as having “none” when no underlying disease was identified. Therefore, an asymptomatic patient could still have underlying medical condition unrelated to the evaluated symptoms.

Regarding the representativeness of the sample pool, we agree with the reviewer that our population is not fully representative of the general pediatric population. Because recruitment took place exclusively at a tertiary-care pediatric referral hospital, these children inherently present a higher prevalence of complex underlying comorbidities and chronic conditions compared to the community.

Therefore, we have explicitly acknowledged this limitation in the "Discussion" section, advising that the findings must be interpreted strictly within the context of this hospital-based cohort.

We added the following information in the “Materials and Methods” section, clarifying the selection criteria:

Lines 108-113, Page 4: “... . Children were enrolled from both patient and outpatient services in the Pediatric Department of the HC-FMRP, USP, irrespective of the presence of respiratory symptoms. Outpatients primarily consisted of children attending scheduled routine consultations and specialized clinical follow-up visits. Inpatients were hospitalized for a variety of acute medical or surgical conditions unrelated to the objectives of this study. ...”

We added the following information in the “Results” section:

Lines 339-344, Page 11: “... In this study, “asymptomatic”, refers to the absence of the clinical symptoms included in the classification (e.g. respiratory, gastrointestinal, neurological or other recorded symptoms) at the time of clinical evaluation. This classification was independent of the presence of underlying medical conditions. Therefore, children classified as asymptomatic could still have previously diagnosed chronic diseases or other medical conditions under clinical follow-up. ...”

We also added an explanation in the “Discussion” section stating that this group could not be compared to the general population.

Lines 477-483, Page 15: “... Because recruitment was conducted exclusively at a tertiary-care pediatric referral hospital, the enrolled children are more likely to present complex underlying comorbidities, chronic medical conditions, and severe clinical manifestations than children in the general community. Consequently, the degree of selection bias is inherent to the design of the study, and the findings should be interpreted within the context of a hospital-based cohort rather than extrapolated directly to the broader pediatric population. ...”

6. L289-293 and Fig 1 A. *Please indicate on the map with arrows the circles representing the patients with MWPYV57 and perhaps the other Genotype 1 isolates with a common symbol since they may represent a local transmission. It would be interesting to see to what extent the positive individuals and the Genotype 1-positived individuals were dispersed among all participants.*

We are grateful for the reviewer's valuable suggestion. Figure 1A has been revised to improve the visualization of the geographic distribution of MWPYV-positive individuals. The sample belonging to

genotype 3 (including MWPvV57) is now highlighted in red, whereas the samples classified taxonomically as genotype 1 s are highlighted in blue. All remaining sampling locations are shown in grey.

Although the clustering of genotype 1 sequences may suggest a possible epidemiological relationship, the phylogenetic analysis did not provide sufficient resolution to support the inference of local transmission. Therefore, we have limited our interpretation to the geographic distribution of the detected genotypes. The revised figure allows a clearer visualization of the spatial dispersion of positive individuals and genotype 1-positive cases among all sampled participants.

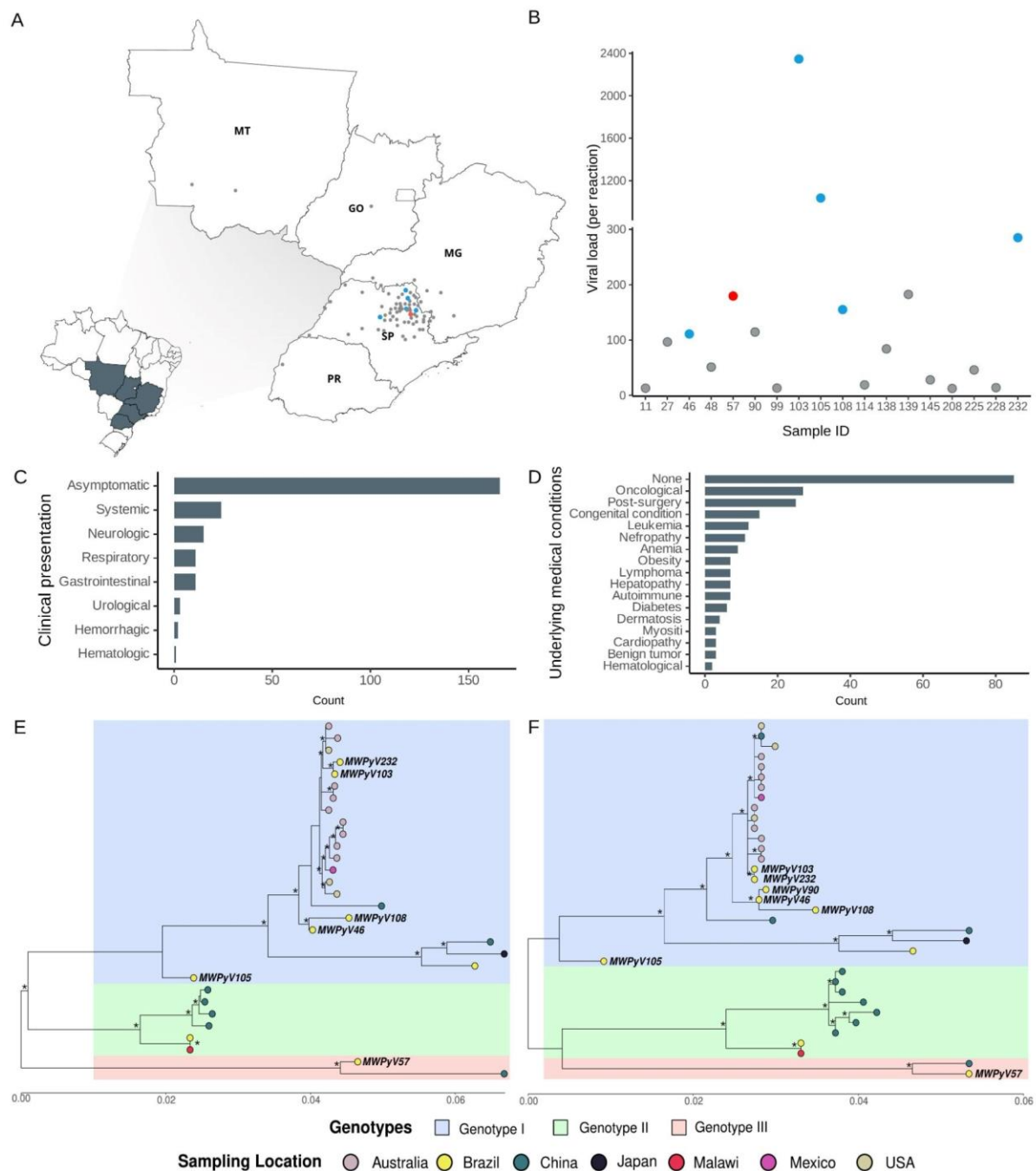


Figure 1. Clinical, geographic, and phylogenetic characterization of MWPvV detected in pediatric patients. A. Geographic origin of patients included in this study. Distribution of pediatric patients by

Brazilian state: Goiás State (GO, n=1), Paraná (PR, n=1), Santa Catarina (SC, n=1), Mato Grosso (MT, n=2), Minas Gerais (MG, n=10), and São Paulo (SP, n=218). Most samples originated from São Paulo State, particularly from Ribeirão Preto and surrounding municipalities, consistent with the recruiting center's location. Sampling sites for which no genotype sequence was obtained are shown in gray. Samples with newly generated complete genomes are highlighted, with the single genotype III isolate (MWPyV57) indicated in red, and the five genotype I isolates indicated in blue. B. Viral load detected in positive samples (n=18). Colors indicate the genotype associated with each sample, as shown in panel A. The y-axis includes a scale break to accommodate high viral load values while improving the visualization of the majority of samples. C. Clinical presentations of the pediatric patients included in this survey. D. Underlying medical conditions of the total number of patients included in this study. E. Whole-genome maximum-likelihood phylogeny of MWPyV genomes, showing the clustering of the genomes into three major genotypes along with the complete genomes that were included in this study. Branches are drawn to scale according to the number of substitutions per site. Ultrafast bootstrap support values >70% are shown on the nodes. Ultrafast bootstrap support values are shown for key basal nodes. Tip colors indicate sampling location. F. Maximum-likelihood phylogeny inferred from VP1 nucleotide sequences of MWPyV along with the sequences generated in this study, showing the three major genotypes (I-III) in accordance with the topology obtained for the complete MWPyV genome. Branches are drawn to scale according to the number of substitutions per site. Ultrafast bootstrap support values >70% are shown on the nodes. Ultrafast bootstrap support values are shown for key basal nodes. Tip colors indicate sampling location.

7. L415-422. Remember polyoma viruses are circular so that regions 1 and 3 are continuous. Perhaps, label the regions in Figure 3A since in your analysis they are not actually continuous since there are excluded sites between them in your analysis.

We thank the reviewer for this helpful comment. Recombination analysis was performed in RDP5 using complete polyomavirus genome sequences, explicitly considering their circular genomic architecture (Line 222), and the software identifies the recombination breakpoints accordingly. Therefore, although the circular nature of the viral genome was accounted for during the RDP5 analysis, we acknowledge that the initial division of the genome into three regions was incorrect. The segments previously designated as regions 1 and 3 are contiguous across the genome origin and should be considered a single genomic partition. Accordingly, only two regions are defined: Region 1, comprising the continuous non-recombinant genomic sequence surrounding the origin, and Region 2, corresponding to the putative recombinant fragment at alignment positions 3450–3842. This has been corrected in the methodological section and in Figure 2 (now), which now includes a circular representation of the MWPyV genome and highlights the inferred recombinant fragment and its corresponding breakpoints. The phylogenetic analysis has also been simplified to include only the two trees reconstructed from these recombinant and non-recombinant genomic partitions, as they fully represent the inferred recombination event. These revisions improve the visualization of the results but do not alter the recombination analysis or its conclusion that the recombinant genome originated from an event involving genotypes 2 and 3.

p-L459-469; L474 ; l539-540 Authors observed low titers in oropharyngeal tissue. Can you assume limited shedding. What is happening in blood, stools and urine which weren't measured? "respiratory origin" the oropharynx is part of both the upper respiratory tract and the upper gastrointestinal tract. In the example of poliovirus infections there is a very short transitory infection of the oropharyngeal cavity followed by viremia and then large-scale excretion if infection spreads to the gastrointestinal tract. Polyoma virus is activated in kidney transplant patients and titers are much higher than in blood. Were nasal swabs considered?

We thank the reviewer for the comment. We completely agree that the detection of MWPvV DNA solely in oropharyngeal swabs alone does not allow providing definitive conclusions regarding the overall dynamics of viral replication, dissemination or shedding in other body compartments.

In the present study, sample collection was restricted to oropharyngeal swabs, and no blood, stool, urine or nasal specimens were available for this analysis. Therefore, we cannot determine whether there was a difference in the viral loads in other anatomical sites or whether MWPvV can be systemically disseminated during infection.

Our interpretation of limited viral shedding was based exclusively on low MWPvV loads detected in oropharyngeal samples analyzed in this study. Accordingly, we have revised our “Discussion” section to clarify that this observation applies specifically to the oropharyngeal compartment and should not be extrapolated to other tissues or biological fluids.

We agree that the oropharynx is anatomically connected to both respiratory and gastrointestinal systems and the biological significance of MWPvV detection specifically in this compartment remains largely uncertain. Different previous studies have reported MWPvV detection in respiratory secretions, fecal samples, and tonsillar tissues suggesting that this virus may be present in multiple anatomical compartments. Additional longitudinal studies, which incorporate multiple specimen types will be required to better define MWPvV tissue tropism, shedding dynamics and transmission routes. We did not include testing of nasal swabs in the study design, and therefore, no comparisons between nasal and oropharyngeal viral loads could be performed.

We added the following information in the “Discussion section”; **Lines 487-495, Page 15:** “... Low-viral loads may also indicate limited viral shedding and inefficient transmission beyond early-life exposure. Additionally, once we investigated only the presence of viral DNA in one body compartment we cannot determine if there were differences in the expression levels in other anatomical sites or if MWPvV can be systematically disseminated during infection. Different previous studies have reported MWPvV detection in respiratory secretions, fecal samples and tonsillar tissues suggesting that this virus might be present in different anatomical sites. Additional longitudinal studies, incorporating multiple types of specimens will be required to better define MWPvV tissue tropism, shedding dynamics and transmission routes. ...”

9. In Figure 3A, one of the excluded areas is the non-coding regulatory region (NCCR). This region undergoes major recombination in both BKV and JCV when replication is enhanced during activation or re-activation. Did you have any indication of any recombination in the NCCR of any of your isolates?

We thank the reviewer for this important comment. We would like to clarify that the NCCR was not intentionally excluded from the recombination analysis. In our dataset, the NCCR could not be reliably evaluated because this short region is located at the beginning of the genome and was not successfully sequenced in the generated complete genomes of this study. This genomic segment was particularly challenging to recover using our sequencing approach; therefore, it was not represented in the final sequence alignment used for the recombination analysis. Consequently, we cannot determine whether rearrangement events or other genomic alterations occurred within the NCCR of our isolates.

We agree that the NCCR is biologically relevant in polyomaviruses, particularly because rearrangements, duplications, and deletions in this region have been associated with enhanced replication and reactivation in BKPyV and JCPyV in different groups of patients. Nevertheless, such alterations are distinct from the homologous recombination events investigated in the present study. We

have revised Figure 3A (including the legend) to explicitly state that the NCCR was not available in our sequence dataset and, therefore, could not be assessed in the recombination analysis.

*Specific comments [Ln is line number “n”]*719110

1. L2 Please add “in Brazil” to the end of the title.

We are grateful to the reviewer for the suggestion. We added “in Brazil” at the end of the title.

2. L46; l481-483. Please explain “...the largest MWPyV genomic data set.” Authors provide WGS data for 6 of their isolates in figure 1D where there are an additional 24 whole genome sequences from other sources (see fig 1D and L174). One of these sources reported 9 whole genome sequences.

We thank the reviewer for pointing this out. We agree that our original wording was inaccurate. Our intention was to emphasize that this study substantially expands the currently available genomic data and provides one of the largest collections of MWPyV genomes from a single study and from a single hospital, rather than to imply that it represents the largest whole genome dataset overall. We have removed this statement from the manuscript.

3. L109-111 “Oropharyngeal swabs were suspended in 2 mL of 1X sterile PBS and homogenized by vortexing prior to aliquoting for nucleic acid extraction.” Suggest replacing “homogenized by vortexing” with “vortexed”

We performed the suggested modification.

4. L209-212. State that the comparisons with BKV and JCV are only in the figures in the supplement as in L371.

We are grateful for the comment of the reviewer. We have revised the “Materials and Methods” section to clarify that BKPyV and JCPyV were included exclusively as outgroups for taxonomic context and were presented only in Supplementary figure 4 and not in the main phylogenetic analysis shown in Figure 1.

Lines 551-552, Page 17: “... BKPyV and JCPyV were included as outgroups to provide taxonomic context. ...”

5. L226 be more specific for “...multiple algorithms.”. Which ones

We thank the reviewer for this suggestion. Although the recombination detection methods are described in the “Materials and Methods” section, we have revised the “Results” section to explicitly identify the algorithms supporting the recombination signal for improved clarity.

We added the following explanation: **Lines 230-231, Page 7: “... Putative recombination breakpoints were inferred based on concordant signals detected by multiple algorithms (RDP, GENECOV, Bootscan and 3Seq). ...”**

6. L393-396; 1496 correct “...distinct lineages...” in the next paragraph you clearly show that the isolate is a mosaic of two genotypes (2 and 3) with the majority of the nucleotides from Genotype 3, hence its position within genotype 3 on the figure.

We thank the reviewer for this observation. We agree that the original wording was imprecise. The recombination analysis indicates the genome of sample MWPYV57 is a mosaic genome that is derived from genotype II and genotype III parenteral lineages, with the majority of the genome corresponding to genotype III. We have revised the text accordingly to explicitly identify the inferred parenteral genotypes and to more accurately describe the phylogenetic placement of the genome of sample MWPYV57.

7. Figures 4 and 5 and legends to figures. Please indicate which sub-figures (A to D) belong to which viral isolate, i.e., Figure 4A and 4B are MWPYV57, etc.

We have identified in the figure legends the isolates, where viral particles were observed.

8. L450-451 “... in oropharyngeal swabs of Brazilian pediatric...”

We have revised the text accordingly.

9. L455-457 How does month of isolation relate to age-dependent increase in seroprevalence?

We thank the reviewer for the identification of this inconsistency. We agree that the temporal distribution of detection across months is not directly related to age-dependent increase in the seroprevalence. The original sentence inadvertently linked to independent observations. We have revised the text to separately describe the temporal distribution of MWPYV detections and the predominance of the infection in very young children, which is consistent with previous serological studies showing age-dependent increase in the MWPYV prevalence.

We have added the following information in the text

Lines 473-475, Page 15: “... Detection occurred sporadically across multiple months, with peaks in August, March, and July, suggesting no clear seasonal pattern. The predominance of MWPYV detection in very young children is also consistent with age-dependent increase in MWPYV seroprevalence³⁵...”

10. KL471 “MWPYV detection was predominantly observed in oropharyngeal cells of asymptomatic children “

We thank the reviewer for this suggestion. We agree that the wording should more precisely reflect the type of samples analyzed. However, our study was based on oropharyngeal swab samples, and therefore we could not determine whether MWPYV was actively infecting oropharyngeal cells, nor can we identify the cellular source of the detected viral DNA. Accordingly, we have revised the text to state that MWPYV DNA was predominantly detected in oropharyngeal swab samples obtained from pediatric patients at the time of sampling, avoiding assumptions regarding cellular tropism or active infection in oropharyngeal mucosa. We added in the text that the detection of MWPYV was performed in “oropharyngeal swab samples” (**Line 324, Page 10**).

11. L487-488 *The genetic similarity to other Deltapolyomaviruses confirmed the membership of the isolates in this group*

We thank the reviewer for this comment. We agree that the placement of MWPyV within the genus *Deltapolyomavirus* is primarily supported by its genetic similarity and phylogenetic relationships to other MWPyV sequences rather by the solely inclusion of BKPyV and JCPyV as outgroups. We have revised the text accordingly to clarify that the outgroups were included to provide taxonomic context and to demonstrate clear separation of MWPyV from the other polyomaviruses. We added the following explanation in the manuscript:

Lines 549-552, Page 17: "... The phylogenetic clustering and genetic similarity of the obtained sequences with previously characterized MWPyV complete genomes confirmed their classification within the Deltapolyomavirus genus. BKPyV and JCPyV were included as outgroups to provide taxonomic context. ..."

12. L522; L525-527 *Why "ancient"? The recombination probably occurred before the specific individual was infected since there is no evidence of the rest of the genotype 2 sequence. However, there is no evidence when the recombination event occurred elsewhere.*

We thank the reviewer for this important observation. We agree that the term "ancient" was not adequately supported by the available data. Our analyses indicate that the recombinant genome was probably present prior to infection of the sampled individual; however they do not allow estimation of when this recombination event occurred in the evolutionary history of MWPyV. Due to the limited availability of complete MWPyV genomes in the public databases, the evolutionary position of this recombinant genome remains uncertain. It is currently unclear whether it represents an intermediate lineage within genotype 3 or a separate genotype altogether. Consequently, additional complete MWPyV genomes are required to clarify its phylogenetic placement and genetic classification. In accordance, we have revised the text to remove the term "ancient" and to avoid implying a specific temporal framework for the inferred recombination event. The following modification was performed in the revised version of the manuscript:

Lines 579-582, Page 18: "... The detection of a genome exhibiting a pattern compatible with genetic exchange between MWPyV genotypes II and III suggests that these genotypes may not be entirely evolutionarily isolated and that recombination events can occur among divergent polyomavirus lineages³⁸. ..."

13. L697 *I was unable to see any red highlights.*

We are sorry for this. We added red color for the MWPyV genotype III.

14. L735 *which "symbols"?*

We inserted three types of symbols: asterisk, circle and square identifying major and minor parenteral lineages.