

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- | | | |
|-------------------------------------|-------------------------------------|--|
| ☐ | ☒ | The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement |
| ☐ | ☒ | A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly |
| ☐ | ☒ | The statistical test(s) used AND whether they are one- or two-sided
<i>Only common tests should be described solely by name; describe more complex techniques in the Methods section.</i> |
| ☐ | ☒ | A description of all covariates tested |
| ☒ | ☐ | A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons |
| ☐ | ☒ | A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals) |
| ☐ | ☒ | For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
<i>Give P values as exact values whenever suitable.</i> |
| ☒ | ☐ | For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings |
| ☒ | ☐ | For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes |
| ☐ | ☒ | Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated |

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection Sequencing was performed on Iseq 100 instrument and the raw data was captured using Iseq System Software v.3. 1.

Data analysis The scripts used for bioinformatic analysis were deposited at the following links: https://github.com/gmcampos27/MWPyV_Analyses and <https://github.com/vcarius/EARLY-LIFE-LOW-LEVEL-INFECTION-AND-GENETIC-DIVERSITY-OF-MALAWI-POLYOMAVIRUS-IN-PEDIATRIC-PATIENTS.git>. The R script used to generate the Brazil collection map and edit the shapefiles is publicly available at https://github.com/jamessiqueirap/scripts_papers/blob/main/MW/MW_collectionMap_generation.R.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our [policy](#)

The assembled genomes were deposited in the NCBI GenBank under the following accession numbers PZ233414-PZ233420 and also available under the following

link <https://doi.org/10.5281/zenodo.19615737>. Additionally, the raw sequencing data were deposited under the BioProject number PRJNA1453041 and biosample accession numbers SAMN57252775, SAMN57252776, SAMN57252777, SAMN7252778, SAMN7252779, SAMN7252780, SAMN7252781.

Research involving human participants, their data, or biological material

Policy information about studies with [human participants or human data](#). See also policy information about [sex, gender \(identity/presentation\), and sexual orientation](#) and [race, ethnicity and racism](#).

Reporting on sex and gender	Sex was recorded for all pediatric participants and included as biological variable in the analysis. Association between detection of MW polyomavirus DNA (positive vs. negative) and demographic and clinical variables, including sex, were evaluated using appropriate statistical methods (Fisher's exact test). No gender related variables were assessed as this information was not available in the medical records.
Reporting on race, ethnicity, or other socially relevant groupings	Information on skin color was extracted from medical records but not included as demographic variable in the statistical analysis. No ethnicity data were available and, and race was not further categorized beyond the classifications present in the clinical records.
Population characteristics	The study included 233 pediatric participants who were aged between 1 month and 18 years (mean 7.35 years and median 6.55 years of age) with both sexes represented (44.6% female and 55.4% male participants). Information on reported race/ethnicity was available and based on this patients were categorized as white (75.5%), mixed race (20.6%) and black (3.9%), according to the medical records. Participants were recruited primarily from the State of Sao Paulo, Brazil with the majority of samples originating from the region of the city of Ribeirao Preto. The study group was clinically highly heterogeneous, including individuals undergoing routine outpatient care as well as patients with diverse clinical conditions ranging from diagnostic investigation to severe immunosuppression. Clinical and laboratory data were extracted from the medical records and used to classify participants according to the clinical status at the time of sampling. No genotypic information or host stratification was available for this study.
Recruitment	Patients were recruited at the Clinic of Pediatrics, University Hospital of the Faculty of Medicine of Ribeirao Preto, (HC-FMRP), University of Sao Paulo during routine clinical visits or while attending outpatient services. Patients (or their legal guardians) were invited to participate, without prior selection based on specific clinical criteria. This type of recruitment strategy may introduce convenience sampling bias, as inclusion depended on patient availability during the study period, was performed only in hospital conditions and may therefore not fully represent the broader pediatric population. Additionally the recruitment presented heterogeneous clinical context (including routine consultations to hospitalized patients with severe clinical conditions) which may have influenced the distribution of the clinical characteristics within this clinical cohort. However, all eligible individuals during recruitment were approached, which helps to exclude selective inclusion at the investigator level.
Ethics oversight	This study was approved by the Institutional Ethics Committee at the University Hospital of the Faculty of Medicine of Ribeirao Preto, University of Sao Paulo (approval number 66508422.0.0000.5440). Written informed consent was obtained from the legal guardians, with assent obtained from the pediatric patients where available. All procedures were conducted in accordance with applicable ethical guidelines involving human participants.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	No a priori sample size calculation was performed. The sample size was determined by the number of pediatric patients available during the study period and who met the inclusion criteria. A post hoc statistical power analysis was conducted using G*power software (v. 3.1.9.7.) to assess, whether the final sample size (n=233) was sufficient. The analysis indicated that the study achieved adequate statistical power (81.73%) with effect size 0.09 at significance level 0.05. The obtained results support that the sample size was sufficient to identify clinically relevant associations between detection of MW Polyomavirus and age or clinical variables.
Data exclusions	No data was excluded from the analysis. All available samples and associated clinical data were included. One duplicate sample was identified and removed to avoid repeated measurements from the same individual. No other exclusions were applied.
Replication	To ensure reproducibility, the extracted respiratory swabs were tested initially once for MWPyV DNA and subsequently reanalyzed in an independent round of screening. Concordant results were obtained across both analyses, with all positive samples consistently identified. In addition, MWPyV-positive samples were subjected to viral load quantification, during which the amplification was successful, further supporting the reliability and reproducibility of the findings. No inconsistencies were observed between the experimental replicates.
Randomization	No randomization was performed as this was an observational study. Participants were not allocated into predefined experimental groups. Instead samples were classified retrospectively based on their MWPyV DNA detection status (positive vs. negative) along with clinical and

demographic data. Statistical analysis were conducted to evaluate associations between variables, which helps account for potential confounding without experimental allocation.

Blinding

Blinding was not performed in this study, as it required access to clinical and demographic data to assess associations with MWPyV infection status. Therefore, investigators were aware of sample identity. All data were handled in accordance with ethical standards, and no sensitive personal identifiers were used in the analysis.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Clinical data
☒	☐ Dual use research of concern
☒	☐ Plants

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Plants

Seed stocks

Report on the source of all seed stocks or other plant material used. If applicable, state the seed stock centre and catalogue number. If plant specimens were collected from the field, describe the collection location, date and sampling procedures.

Novel plant genotypes

Describe the methods by which all novel plant genotypes were produced. This includes those generated by transgenic approaches, gene editing, chemical/radiation-based mutagenesis and hybridization. For transgenic lines, describe the transformation method, the number of independent lines analyzed and the generation upon which experiments were performed. For gene-edited lines, describe the editor used, the endogenous sequence targeted for editing, the targeting guide RNA sequence (if applicable) and how the editor was applied.

Authentication

Describe any authentication procedures for each seed stock used or novel genotype generated. Describe any experiments used to assess the effect of a mutation and, where applicable, how potential secondary effects (e.g. second site T-DNA insertions, mosaicism, off-target gene editing) were examined.