nature portfolio

Peer Review File

Epidemiological characteristics of monkeypox virus Clade Ib in the Democratic Republic of the Congo

Corresponding Author: Ms Cécile Kremer

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 titled “Epidemiological characteristics of mpox virus Clade Ib in the Democratic Republic of the Congo: the impact of transmission mode” addresses an important public health issue and presents novel epidemiological data on MPXV Clade Ib. However, numerous methodological, ethical, and reporting issues must be addressed before the manuscript can be considered for publication. The authors should thoroughly revise the text, clarify data inconsistencies, justify key methodological choices, and ensure full transparency in the presentation of results. Only after these substantial revisions can the scientific merit of the work be properly evaluated.

Title and Line 41 - Incorrect virus nomenclature: "monkeypox virus" vs. "mpox virus". It may be advisable to revise the term 'mpox virus,' as it is not aligned with current WHO and ICTV nomenclature guidelines. While 'mpox' is now the recommended term for the disease, the virus name remains monkeypox virus (MPXV) and should be used as such in scientific contexts to ensure clarity and consistency.
Line 97 - Although the authors state that 1013 individuals tested PCR-positive for mpox, it would be more accurate to clarify that PCR detects the presence of monkeypox virus (MPXV) DNA, not the disease itself. Given this distinction, the authors should consider rephrasing the sentence to reflect that PCR positivity indicates viral detection, which may or may not correspond to clinical mpox.
Line 103 - The total denominator for this category should be corrected to 578, as currently stated values (e.g., 'farmer (40/1013, 6.92%)') are inconsistent. Please revise the percentage and denominator accordingly to ensure internal consistency.
Lines 99 and 106 - The manuscript reports that 475 out of 657 MPXV-infected individuals over 12 years of age engaged in sexual intercourse in the preceding three weeks, including 162 who reported sexual contact with sex workers. While the ethical section clarifies that participants aged 12–17 provided assent and parental consent, the inclusion of minors in behavioural analyses, particularly on sensitive topics such as sexual activity, raises important ethical and interpretative concerns. This is especially relevant given that the referenced Kamituga study defined adults as individuals aged 15 years or older. The authors should clearly specify the age thresholds used in these analyses, indicate whether data from younger adolescents (e.g., 12–14 years) were included, and provide justification for their inclusion. Transparent reporting and careful contextualisation of these data are essential to avoid misinterpretation, even when appropriate ethical safeguards are in place.
Line 114 – The presentation of numbers and percentages throughout the manuscript lacks consistency. In some instances, the authors report values as 'X/Y (Z%)', while in others only 'X (Z%)' is presented, without indicating the denominator. This inconsistency makes it difficult for readers to interpret and compare the data accurately. It is recommended that the authors adopt a uniform format, preferably 'X/Y (Z%)', to enhance clarity and facilitate interpretation across the text.
Line 121 - There appears to be a discrepancy in the reported numbers of participants. The manuscript states that the Kamituga and Goma cohorts included 650 and 313 participants, respectively, totalling 963 individuals. However, it later refers to '1000 MPXV-infected individuals from Kamituga and Goma for whom the date of symptom onset was available.' This figure exceeds the combined cohort size and is not explained. The authors should clarify whether this number includes additional participants not previously mentioned, reflects a rounding approximation, or is a typographical error.
Line 140 - Throughout the manuscript, numbers below ten should be written out in full (e.g., 'three' instead of '3') in accordance with standard scientific writing conventions, unless they refer to units of measurement, statistical values, or figure/table labels.
Line 141 - In the section describing the distribution of sexual and non-sexual contacts, the authors present absolute numbers

followed by percentages for each subgroup. However, in several cases, the percentages are only provided for the subcategories, and not for the total counts mentioned at the beginning of each sentence (e.g., 'Among 48 reported sexual contacts…'). For clarity and consistency, it is recommended that the authors also indicate the percentage that each total (e.g., 48, 40, 23) represents within the broader dataset. This would allow readers to better understand the relative weight of each subgroup in the overall analysis.
Line 152 - The manuscript states that 157 MPXV-infected individuals with known symptom onset reported contact with at least one other study participant, resulting in 260 individuals being included in the serial interval estimation. However, the total number of participants with known symptom onset is not provided.
Lines 152 to 161 – See comment of line 114.
Line 182 – The text refers to 'OPX-Ct' values, while the Methods section clearly states that the molecular diagnosis specifically targeted MPXV, not the broader Orthopoxvirus (OPXV) genus. This discrepancy should be addressed, and the terminology in the text revised if necessary to ensure consistency. Additionally, a brief explanation of Ct values would be helpful—for example, clarifying that lower Ct values correspond to higher viral loads. See also comment Supplementary Line 19, regarding Figure S2.
Line 182 - It should also be noted that the type of clinical samples used for PCR diagnosis is only mentioned in this paragraph; for transparency and reproducibility, this information should be clearly stated in the Methods section as well. See also comment Line 269.
Lines 205 and 207 – The authors should include the specific values for Uvira, DRC, to ensure completeness and allow for proper comparison across study sites.
Line 219 - This article focuses on the epidemiological characteristics of Clade Ib of the monkeypox virus in the Democratic Republic of the Congo. However, in the Discussion section, the authors state that clade-specific PCR or sequencing was not performed, which limits their ability to confirm that all detected cases belong to Clade Ib. Despite the statement that “the epidemiological context strongly suggests Clade Ib circulation in both study sites”, this assumption is only partially supported by available evidence. While sequencing data published up to the end of 2024 confirm Clade Ib as the predominant clade in Goma and Kamituga, Clade Ia is also known to be circulating in other regions of the DRC. Moreover, clade information is available for only 253 of the 1,013 cases in the study, representing just 25% of the total sample. This limits the representativeness of the findings. In addition, many mpox cases in the DRC are not tested or confirmed by PCR due to limited diagnostic capacity, meaning that WHO-reported data may not fully reflect the actual clade distribution or extent of circulation. To strengthen their conclusions, the authors should clarify the basis for assuming exclusive Clade Ib circulation and, where possible, expand clade determination across a more representative sample.
Line 265 - The manuscript states that questions about sexual behaviour were restricted to individuals over 12 years old. As noted in previous comments, this raises important ethical and interpretative concerns, particularly when minors are involved. The authors should provide a clear and detailed justification for selecting this age threshold, including any legal, cultural, or methodological rationale that supports its use in the study context.
Line 269 - The manuscript describes the use of the GeneXpert® Mpox PCR platform for diagnostic testing but does not specify the type of clinical samples (e.g., lesion swabs, oropharyngeal swabs) used for PCR analysis. Given that the type of specimen can influence diagnostic sensitivity and interpretation of results, the authors should clearly state which sample types were collected and tested.
Line 281 - The manuscript would benefit from a brief description of the statistical models or algorithms used, including the purpose of each within the overall analysis. While some methods are mentioned (e.g., Bayesian regression), it is not always clear what specific research question each model addresses. To improve clarity and transparency, the authors should also comment on the advantages and limitations of the chosen models compared to alternative approaches.
Line 304 - The acronym MCMC (Markov Chain Monte Carlo?) is used in the manuscript, but no definition or explanation is provided at first mention. For clarity, especially for readers who may not be familiar with this statistical method, the full term should be spelled out and briefly described when it first appears.
Line 380 - There is a typographical error in the word 'dynamics' that should be corrected.
Line 426 - In Tables 1, 2, and 3, as well as in Figure 1, S2 and S3, the sample size (n) for each group or scenario is not indicated. This contrasts with Tables 4 and 5, where the number of observations is clearly specified for each subgroup. For clarity and transparency, it is recommended that the authors include the corresponding sample sizes in all tables and figures presenting stratified analyses.
Supplementary
Line 16 - Supplementary Figure S1 illustrates the age distribution of 1,013 MPXV-infected individuals; however, several aspects of the figure should be improved for clarity and consistency. First, the leftmost bar (representing the youngest age group, 0–5 years) appears misaligned with the Y-axis, giving a misleading visual impression. This should be corrected to ensure proper alignment. It should also be clarified whether the Y-axis represents the absolute number of cases or frequency (relative proportion), as this distinction is important for interpreting the distribution accurately. Second, the X-axis scale should be adjusted to reflect the age categories explicitly used in the manuscript (≤12 years, 12–17 years, and ≥18 years), which would allow for better visual interpretation in line with the text. Additionally, the X-axis label should specify the unit of measurement (i.e., 'Age (years)').
Line 19 - Supplementary Figure S2 provides useful comparisons of Ct values across study sites, transmission modes, age groups, and timing of sample collection. However, several improvements would enhance clarity. First, the figure refers to 'OPX-Ct' values, but in the Methods section, it is stated that molecular diagnosis targeted MPXV specifically, not generic Orthopoxvirus (OPXV). This discrepancy should be clarified, and the labelling in the figure adjusted if necessary. Additionally, the meaning of Ct values should be explained (e.g., lower Ct values indicating higher viral load), and the colour coding for transmission mode (sexual vs. non-sexual) should be made more explicit through a clear legend. The age groups on the X-axis are currently labelled numerically (1, 2, 3), which requires the reader to reference a footnote; direct labelling (e.g., 'Adults', '12–17 y', '<12 y') would improve readability. Lastly, the font size throughout the figure is too small and should be increased for clarity.
Line 23 - Supplementary Figure S3 shows trace plots from the Bayesian model used to estimate the serial interval. The plots

appear stable and consistent, suggesting that the model worked well. However, the figure would benefit from clearer labelling: each graph should indicate which parameter it refers to, and the axes could be better defined. This would make the figure easier to understand, especially for readers who are not specialists in Bayesian analysis.

(Remarks on code availability)

Reviewer #2

(Remarks to the Author)
This study provides valuable insights into how key epidemiological parameters—such as the serial interval and incubation period, differ depending on the mode of transmission in the context of MPXV Clade Ib. These findings are clearly important for informing public health responses and improving outbreak modeling.

At the same time, I would encourage the authors to further articulate the broader scientific implications of these findings, particularly for audiences beyond the field of infectious disease epidemiology. Clarifying how this work contributes to more general scientific understanding, for example, regarding transmission dynamics, host-pathogen interactions, or methodological approaches, could help enhance the relevance and impact of the manuscript across disciplines.

minor comments
L155 and 177, I would appreciate more clarity on how "evidence for a difference" was determined or defined in this context.
Figure, The lines in the figure are thin and difficult to distinguish, making it hard to tell which corresponds to sexual or household transmission.

(Remarks on code availability)

Reviewer #3

(Remarks to the Author)
This is an interesting study on how transmission mode may influence serial intervals and incubation periods in mpox virus transmission. I mainly have questions on the how the data were collected and possible biases.

• Line 115: how come that so many contacts were participants of the study? Were they recruited into the study?
• Would it be possible to add a flow diagram or table showing the various percentages of participants in different categories or used for specific analyses? That would be very helpful for the reader.
• Lines 127-129: What is the rationale for choosing participants who reported only one contact with another participant? Is that not going to introduce a bias into the estimates? Would it be possible to average over several contacts? Also, could these participants have had other contacts outside the study population?
• Line 140: You call these pairs transmission pairs, but this relies only on selfreported information about contacts, or was there some way of confirming transmission?
• Figure 1: It is unclear to me whether the transmission network is based somehow on the assumption that the population is closed, or how it takes into account that contacts can be with persons, who are not participants of the study.
• Line 258: Please give some more background information about this study site. What is the population size?
• Lines 266-268: Did the participants name their contacts? Or how could investigators know whether they were study participants?
• There should be more discussion of any possible biases in the data, for example sampling bias
• Line 306: Please give some information about what you assume about transmission trees. How are they defined?

(Remarks on code availability)
I did not have time to check the code and reproduce the results. There is a Readme file, which states which scripts are included in the repository.

Version 1:

Reviewer comments:

Reviewer #1

(Remarks to the Author)
Dear Editor,

Thank you for the opportunity to review the revised version of the manuscript titled “Epidemiological characteristics of monkeypox virus Clade Ib in the Democratic Republic of the Congo: the impact of transmission mode”.

After carefully reviewing both the rebuttal letter and the revised manuscript, I confirm that the authors have addressed the reviewers’ comments in a thorough and appropriate manner. The manuscript has been significantly improved in terms of clarity, contextual interpretation, and scientific depth.

I therefore recommend this manuscript for publication.

Thank you very much.

Best regards,

(Remarks on code availability)

REVIEWER COMMENTS

Reviewer #1 (Remarks to the Author):

The manuscript titled “Epidemiological characteristics of mpox virus Clade Ib in the Democratic Republic of the Congo: the impact of transmission mode” addresses an important public health issue and presents novel epidemiological data on MPXV Clade Ib. However, numerous methodological, ethical, and reporting issues must be addressed before the manuscript can be considered for publication. The authors should thoroughly revise the text, clarify data inconsistencies, justify key methodological choices, and ensure full transparency in the presentation of results. Only after these substantial revisions can the scientific merit of the work be properly evaluated.

We thank the reviewer for recognizing the importance of our manuscript and their thorough review. We have revised the manuscript addressing all the issues raised by the reviewer and believe the manuscript is greatly improved by this feedback.

1. Title and Line 41 - Incorrect virus nomenclature: "monkeypox virus" vs. "mpox virus". It may be advisable to revise the term 'mpox virus,' as it is not aligned with current WHO and ICTV nomenclature guidelines. While 'mpox' is now the recommended term for the disease, the virus name remains monkeypox virus (MPXV) and should be used as such in scientific contexts to ensure clarity and consistency.

We thank the reviewer for this comment and have consistently changed the term ‘mpox virus’ to ‘monkeypox virus’ throughout the revised manuscript.

2. Line 97 - Although the authors state that 1013 individuals tested PCR-positive for mpox, it would be more accurate to clarify that PCR detects the presence of monkeypox virus (MPXV) DNA, not the disease itself. Given this distinction, the authors should consider rephrasing the sentence to reflect that PCR positivity indicates viral detection, which may or may not correspond to clinical mpox.

We thank the reviewer for this comment and have adapted the sentence on lines 104-107 in the revised manuscript as follows:

“Between May 2024 and February 2025, 1013 individuals that were suspected of having clinical mpox disease tested PCR-positive for OPXV DNA and negative for MPXV Clade II DNA, indicating viral detection of MPXV Clade I, and were included in the current study.”

In addition, we have added more details about diagnostic testing performed in the Methods section based on the reviewer’s comment 11.

3. Line 103 - The total denominator for this category should be corrected to 578, as currently stated values (e.g., 'farmer (40/1013, 6.92%)') are inconsistent. Please revise the percentage and denominator accordingly to ensure internal consistency.

We apologize for this oversight and have corrected the denominator in the revised manuscript on line 113.

4. Lines 99 and 106 - The manuscript reports that 475 out of 657 MPXV-infected individuals over 12 years of age engaged in sexual intercourse in the preceding three weeks, including 162 who reported sexual contact with sex workers. While the ethical section clarifies that participants aged 12–17 provided assent and parental consent, the inclusion of minors in

behavioural analyses, particularly on sensitive topics such as sexual activity, raises important ethical and interpretative concerns. This is especially relevant given that the referenced Kamituga study defined adults as individuals aged 15 years or older. The authors should clearly specify the age thresholds used in these analyses, indicate whether data from younger adolescents (e.g., 12–14 years) were included, and provide justification for their inclusion. Transparent reporting and careful contextualization of these data are essential to avoid misinterpretation, even when appropriate ethical safeguards are in place.

We agree with the reviewer that transparent reporting of these sensitive data is essential. Although we understand the reviewer's concern about including behavioral data from younger adolescents, understanding transmission dynamics of MPXV Clade Ib requires a comprehensive examination of all age groups, including adolescents. Although sexual activity among minors is a sensitive topic, excluding this information could bias epidemiological estimates or underestimate potential transmission pathways, especially in communities where early sexual debut occurs. As the reviewer recognizes, the data used were collected in compliance with all relevant ethical guidelines, with appropriate IRB or ethics committee approvals in place. Including sexual activity data from younger adolescents acknowledges its occurrence and integrates it into a realistic and data-driven assessment of epidemiological characteristics. The manuscript treats all such data with clinical and scientific objectivity, without sensationalism.

Although Brosius et al. defined adults as those individuals over 15 years of age, the MBOTE study did record behavioral data related to sexual activity for individuals aged 12-17 if parental consent was given and a parent was present during the interview. We have clarified this in the revised manuscript on lines 321-323 as follows:

"Questions about sexual behavior were restricted to individuals from 12 years onward. Although sexual activity among younger adolescents (e.g. 12-14 years old) is a sensitive topic, excluding this information could bias epidemiological estimates."

Of the 475 individuals who engaged in sexual intercourse in the preceding three weeks, only two were children under 15 years of age. Due to this low number, changing the cutoff in the analyses to 15 years did not change the results. We have added the following to the revised manuscript on lines 117-119:

"Among these, only 2/475 (0.42%) were children under 15 years of age, and 154/475 (32.42%) reported sexual contact with a sex worker, of which 1/154 (0.65%) was under 15 years of age."

5. Line 114 – The presentation of numbers and percentages throughout the manuscript lacks consistency. In some instances, the authors report values as 'X/Y (Z%)', while in others only 'X (Z%)' is presented, without indicating the denominator. This inconsistency makes it difficult for readers to interpret and compare the data accurately. It is recommended that the authors adopt a uniform format, preferably 'X/Y (Z%)', to enhance clarity and facilitate interpretation across the text.

We appreciate this comment and have revised the manuscript accordingly, making sure to always include the numerator, denominator, and percentage.

6. Line 121 - There appears to be a discrepancy in the reported numbers of participants. The manuscript states that the Kamituga and Goma cohorts included 650 and 313 participants, respectively, totalling 963 individuals. However, it later refers to '1000 MPXV-infected

individuals from Kamituga and Goma for whom the date of symptom onset was available.' This figure exceeds the combined cohort size and is not explained. The authors should clarify whether this number includes additional participants not previously mentioned, reflects a rounding approximation, or is a typographical error.

We thank the reviewer for noting this and apologize for a typing error on line 117 in the submitted manuscript. Instead of 313, there were 363 participants from Goma included. Together with the participants from Kamituga, this results in 1013 participants in total. Of those 1013 participants, symptom onset data was available for 1000/1013 (98.72%). We have made the necessary corrections at the appropriate locations in the revised manuscript.

7. Line 140 - Throughout the manuscript, numbers below ten should be written out in full (e.g., 'three' instead of '3') in accordance with standard scientific writing conventions, unless they refer to units of measurement, statistical values, or figure/table labels.

We thank the reviewer for noting this and have revised the manuscript accordingly.

8. Line 141 - In the section describing the distribution of sexual and non-sexual contacts, the authors present absolute numbers followed by percentages for each subgroup. However, in several cases, the percentages are only provided for the subcategories, and not for the total counts mentioned at the beginning of each sentence (e.g., 'Among 48 reported sexual contacts…'). For clarity and consistency, it is recommended that the authors also indicate the percentage that each total (e.g., 48, 40, 23) represents within the broader dataset. This would allow readers to better understand the relative weight of each subgroup in the overall analysis.

We thank the reviewer for this comment and have included denominators for all counts in the revised manuscript. We also note that some of these counts have changed due to an oversight regarding one duplicated participant in the serial interval estimation (N = 110 instead of 111), which has been corrected in the revised manuscript.

9. Line 152 - The manuscript states that 157 MPXV-infected individuals with known symptom onset reported contact with at least one other study participant, resulting in 260 individuals being included in the serial interval estimation. However, the total number of participants with known symptom onset is not provided.

We have now reported the total number of participants with a known symptom onset date in the revised manuscript, both overall (line 133) and for the Kamituga site specifically (line 168) as only the data from Kamituga were included in the serial interval estimation due to limited transmission pair data being available from Goma.

10. Lines 152 to 161 – See comment of line 114.

We have now included numerator, denominator, and percentage for all counts in the revised manuscript.

11. Line 182 – The text refers to 'OPX-Ct' values, while the Methods section clearly states that the molecular diagnosis specifically targeted MPXV, not the broader Orthopoxvirus (OPXV) genus. This discrepancy should be addressed, and the terminology in the text revised if necessary to ensure consistency. Additionally, a brief explanation of Ct values would be helpful—for example, clarifying that lower Ct values correspond to higher viral loads. See also comment Supplementary Line 19, regarding Figure S2.

We thank the reviewer for noting this discrepancy. We acknowledge the lack of clarity about PCR testing and have elaborated more on this in the revised manuscript on lines 326-332 as follows:

"Diagnostic testing of skin lesion and, in some cases, oropharyngeal swabs was conducted on site using Xpert® Mpox cartridges on a four-module GeneXpert real-time PCR system (Cepheid, Sunnyvale, CA, USA), which detects non-variola orthopoxvirus (OPXV) and MPXV Clade II DNA. A sample was considered positive for MPXV Clade I if non-variola OPXV DNA was detected with a cycle threshold (Ct) value below 40 and if MPXV Clade II PCR was negative. Clade confirmation with a Clade Ib-specific PCR was performed on a subset of samples collected between 2 May and 30 August 2024."

We have further added a brief explanation of Ct values in the revised manuscript on lines 204-206 as follows:

"Ct values indicate the number of cycles needed to detect the target genetic material, with lower Ct values meaning that higher amounts of the target are present in the sample, and thus reflecting a higher viral load."

12. Line 182 - It should also be noted that the type of clinical samples used for PCR diagnosis is only mentioned in this paragraph; for transparency and reproducibility, this information should be clearly stated in the Methods section as well. See also comment Line 269.

We have now specified in the Methods that skin lesion as well as oropharyngeal samples were used for PCR diagnosis as follows on line 326 in the revised manuscript:

"Diagnostic testing of skin lesion and, in some cases, oropharyngeal swabs was conducted on site using Xpert® Mpox cartridges …"

13. Lines 205 and 207 – The authors should include the specific values for Uvira, DRC, to ensure completeness and allow for proper comparison across study sites.

We thank the reviewer for this comment. The reason we did not initially include the specific incubation period estimates from Uvira was that they reflect the time between infection and rash onset, while we estimate the incubation period as time between infection and the onset of prodromal symptoms, which does not necessarily correspond to rash onset. We have now included the estimates from Uvira in the revised manuscript on lines 229-233 as follows, noting the important distinction between time to symptoms and time to rash:

"In that study, the median incubation period from exposure to first symptoms was 16.1 days (95% CrI 12.8 – 20.1 days), substantially longer than our estimates. However, we estimated a mean incubation period from exposure to first symptoms of 8.51 days (95% CrI 6.84 – 10.54 days) for sexual transmission, which is similar to the previous estimate of a median of 9.5 days (95% CrI 3.7 – 16.4 days) from exposure to rash onset [13]."

14. Line 219 - This article focuses on the epidemiological characteristics of Clade Ib of the monkeypox virus in the Democratic Republic of the Congo. However, in the Discussion section, the authors state that clade-specific PCR or sequencing was not performed, which limits their ability to confirm that all detected cases belong to Clade Ib. Despite the statement that "the epidemiological context strongly suggests Clade Ib circulation in both study sites", this assumption is only partially supported by available evidence. While sequencing data published up to the end of 2024 confirm Clade Ib as the predominant clade in Goma and Kamituga, Clade

Ia is also known to be circulating in other regions of the DRC. Moreover, clade information is available for only 253 of the 1,013 cases in the study, representing just 25% of the total sample. This limits the representativeness of the findings. In addition, many mpox cases in the DRC are not tested or confirmed by PCR due to limited diagnostic capacity, meaning that WHO-reported data may not fully reflect the actual clade distribution or extent of circulation. To strengthen their conclusions, the authors should clarify the basis for assuming exclusive Clade Ib circulation and, where possible, expand clade determination across a more representative sample.

Indeed, clade determination was only performed for a subset of the study participants that were recruited between May and August 2024, and the majority (98%) of those were confirmed as Clade Ib. In addition, all available sequences from North and South Kivu since 2023 have been confirmed as Clabe Ib (see https://worldhealthorg.shinyapps.io/mpx_global/ and https://nextstrain.org/groups/inrb-mpox/clade-I). These include all sequencing performed whether or not the patients were included in the MBOTE study. Furthermore, mpox cases are historically rare in the Kivu provinces and have only been described in areas bordering more forested neighboring provinces, not in the urbanized regions where the current epidemic is unfolding. Unfortunately, it was not possible to expand clade determination for the present study. We have added the following to the revised manuscript on lines 246-250:

"*However, the epidemiological context strongly suggests Clade Ib circulation in both study sites, as all available sequences from North and South Kivu since 2023 have been confirmed as Clade Ib. In addition, in a subset of 253 participants, it was previously confirmed that they were infected with Clade Ib using a lad-specific PCR.*"

Furthermore, of the 147/260 (56.54%) mpox patients recruited between May and August 2024 who were included in the Bayesian regression model for the serial interval, 105/147 (71.43%) were confirmed as Clade Ib. For 3/147 (2.04%), the Clade Ib-specific PCR was negative, possibly because of sample degradation since these also tested negative for Clade Ia. For the remaining 39/147 (26.53%) no clade-specific PCR test was performed.

15. Line 265 - The manuscript states that questions about sexual behaviour were restricted to individuals over 12 years old. As noted in previous comments, this raises important ethical and interpretative concerns, particularly when minors are involved. The authors should provide a clear and detailed justification for selecting this age threshold, including any legal, cultural, or methodological rationale that supports its use in the study context.

As written in our reply to comment 4 above, we have included the following justification in the revised manuscript on lines 322-323:

"Although sexual activity among younger adolescents (e.g. 12-14 years old) is a sensitive topic, excluding this information could bias epidemiological estimates."

16. Line 269 - The manuscript describes the use of the GeneXpert® Mpox PCR platform for diagnostic testing but does not specify the type of clinical samples (e.g., lesion swabs, oropharyngeal swabs) used for PCR analysis. Given that the type of specimen can influence diagnostic sensitivity and interpretation of results, the authors should clearly state which sample types were collected and tested.

We apologize for this oversight and thank the reviewer for noting this. Diagnostic testing was performed on skin lesion and, in some cases, oropharyngeal swabs. We have included this in the revised manuscript on lines 326-332 as follows:

"Diagnostic testing of skin lesion and, in some cases, oropharyngeal swabs was conducted on site using Xpert® Mpox cartridges on a four-module GeneXpert real-time PCR system (Cepheid, Sunnyvale, CA, USA), …"

17. Line 281 - The manuscript would benefit from a brief description of the statistical models or algorithms used, including the purpose of each within the overall analysis. While some methods are mentioned (e.g., Bayesian regression), it is not always clear what specific research question each model addresses. To improve clarity and transparency, the authors should also comment on the advantages and limitations of the chosen models compared to alternative approaches.

We thank the reviewer for this comment and have elaborated on the statistical models used for estimation of the serial interval in the revised manuscript on lines 347-350 and 3563-365 as follows:

"This approach is fully data-driven and combines a nonparametric estimator of the cumulative distribution function with the bootstrap to yield point and interval estimates of any desired feature of the serial interval distribution [14], in this case the mean and standard deviation."

"In addition, to formally assess differences in serial intervals based on certain characteristics, we used a Bayesian linear regression model to investigate the influence of sexual and household transmission on the mean serial interval, while accounting for age."

For estimation of the incubation period, we have revised the methods on lines 412-414, 429-431, and 437-438 as follows:

"In this approach, adaptive Hamiltonian Monte Carlo (HMC) is used to draw samples from the posterior distribution of the Weibull parameters, enabling full Bayesian inference with uncertainty quantification."

"Again, to formally assess differences in incubation periods based on certain characteristics, we used a Bayesian linear regression model to investigate the influence of sexual transmission and age on the incubation period."

"We used HMC to evaluate the posterior distribution using four chains with 10 000 iterations each, with a burn-in period of 50%."

Regarding the limitations of the chosen models, in particular for estimation of the incubation period in which we had to resort to using relatively wide exposure windows, this is mentioned in the Discussion on lines 268-271. In addition, we have elaborated a bit more on potential biases in the Discussion on lines 258-267 as per Comments 3 and 8 of Reviewer 3.

18. Line 304 - The acronym MCMC (Markov Chain Monte Carlo?) is used in the manuscript, but no definition or explanation is provided at first mention. For clarity, especially for readers who may not be familiar with this statistical method, the full term should be spelled out and briefly described when it first appears.

We thank the reviewer for this comment and have clarified this in the revised manuscript as follows on lines 374-376:

"Markov chain Monte Carlo (MCMC) was used to estimate the parameters of the regression model, in which random sampling is used to explore a range of possible parameter values and assess which values are most likely given the data."

19. Line 380 - There is a typographical error in the word 'dynamics' that should be corrected.

We thank the reviewer for noting this and have corrected this error in the revised manuscript.

20. Line 426 - In Tables 1, 2, and 3, as well as in Figure 1, S2 and S3, the sample size (n) for each group or scenario is not indicated. This contrasts with Tables 4 and 5, where the number of observations is clearly specified for each subgroup. For clarity and transparency, it is recommended that the authors include the corresponding sample sizes in all tables and figures presenting stratified analyses.

We have now indicated the sample size for Figure 1 (N = 260), Supplementary Figure S2 (N = 960), and Table 2 (N = 260). For Supplementary Figure S3, we believe sample size is not relevant as this graph simply provides evidence of MCMC convergence. Regarding Table 1 and Table 3, we believe adding sample sizes for each subgroup would make the tables messy for the readers, therefore we would suggest keeping them as they are. However, we added the sample size used in each analysis in Supplementary Table S1 as per Comment 2 of Reviewer 3.

Supplementary

21. Line 16 - Supplementary Figure S1 illustrates the age distribution of 1,013 MPXV-infected individuals; however, several aspects of the figure should be improved for clarity and consistency. First, the leftmost bar (representing the youngest age group, 0–5 years) appears misaligned with the Y-axis, giving a misleading visual impression. This should be corrected to ensure proper alignment. It should also be clarified whether the Y-axis represents the absolute number of cases or frequency (relative proportion), as this distinction is important for interpreting the distribution accurately. Second, the X-axis scale should be adjusted to reflect the age categories explicitly used in the manuscript (≤12 years, 12–17 years, and ≥18 years), which would allow for better visual interpretation in line with the text. Additionally, the X-axis label should specify the unit of measurement (i.e., 'Age (years)').

We thank the reviewer for this comment. We have adjusted the x-axis to include breaks at each year, with red dashed lines indicating age 12 and 18 specifically. The reviewer is correct that the y-axis represents the absolute number of cases, and we have adjusted the figure label accordingly. We hope that the Figure now gives a better visualization of the age distribution.

22. Line 19 - Supplementary Figure S2 provides useful comparisons of Ct values across study sites, transmission modes, age groups, and timing of sample collection. However, several improvements would enhance clarity. First, the figure refers to 'OPX-Ct' values, but in the Methods section, it is stated that molecular diagnosis targeted MPXV specifically, not generic Orthopoxvirus (OPXV). This discrepancy should be clarified, and the labelling in the figure adjusted if necessary. Additionally, the meaning of Ct values should be explained (e.g., lower Ct values indicating higher viral load), and the colour coding for transmission mode (sexual vs. non-sexual) should be made more explicit through a clear legend. The age groups on the X-axis are currently labelled numerically (1, 2, 3), which requires the reader to reference a

footnote; direct labelling (e.g., 'Adults', '12–17 y', '<12 y') would improve readability. Lastly, the font size throughout the figure is too small and should be increased for clarity.

We acknowledge the lack of clarity about PCR testing and the meaning of Ct values, and have elaborated more on this in the revised manuscript as mentioned in our reply to comment 11 above. We have also revised Supplementary Figure S2 according to the reviewer's suggestions.

23. Line 23 - Supplementary Figure S3 shows trace plots from the Bayesian model used to estimate the serial interval. The plots appear stable and consistent, suggesting that the model worked well. However, the figure would benefit from clearer labelling: each graph should indicate which parameter it refers to, and the axes could be better defined. This would make the figure easier to understand, especially for readers who are not specialists in Bayesian analysis.

We thank the reviewer for this comment and have revised Supplementary Figure 3 according to the reviewer's suggestions.

Reviewer #2 (Remarks to the Author):

This study provides valuable insights into how key epidemiological parameters—such as the serial interval and incubation period, differ depending on the mode of transmission in the context of MPXV Clade Ib. These findings are clearly important for informing public health responses and improving outbreak modeling.

We would like to thank the reviewer for recognizing the importance of our work in informing public health responses.

At the same time, I would encourage the authors to further articulate the broader scientific implications of these findings, particularly for audiences beyond the field of infectious disease epidemiology. Clarifying how this work contributes to more general scientific understanding, for example, regarding transmission dynamics, host-pathogen interactions, or methodological approaches, could help enhance the relevance and impact of the manuscript across disciplines.

We thank the reviewer for this suggestion and have revised the Discussion section to emphasize the broader scientific implications as follows on lines 276-305 in the revised manuscript:

"Differences in key epidemiological characteristics, such as the serial interval and incubation period, between sexual and non-sexual transmission are highly relevant for public health policy. Shorter serial intervals for sexual transmission suggest that infections can propagate more quickly, underscoring the need for rapid public health responses such as expedited case detection, contact tracing, and targeted vaccination in areas where sexual contact is the dominant transmission mode. Accurate and context-specific estimates of these epidemiological characteristics also improve transmission models, enabling better prediction of outbreak dynamics and the impact of targeted public health strategies, which enhances the ability of policymakers across disciplines to allocate resources efficiently. Our findings also suggest that host-pathogen interactions may not be fixed, but modulated by the mode and context of exposure. This aligns with the idea that transmission route can influence pathogen evolution and host response through mechanisms like inoculum size or site-specific viral

replication dynamics, and such within-host dynamics should be taken into account in infectious disease models.

Beyond infectious disease modeling, these differences in epidemiological characteristics have practical implications for public health policy and planning. For example, quarantine durations and isolation guidelines should be tailored to the predominant transmission mode in a given setting. Failure to account for these differences could lead to either unnecessarily burdensome restrictions or insufficient containment of the outbreak. Although MPXV Clade Ib was first thought to be transmitted mainly through sexual contact, the shift in age of affected individuals back to younger individuals in other areas indicates that the context of the outbreak setting is an essential factor to consider when devising public health policy [16]. In light of this, there is a need for systematic data collection in different areas in order to characterize the changing epidemiology of MPXV. Ultimately, these findings underscore the need for flexible, data-driven public health strategies that account for both biological and social determinants of transmission."

minor comments

1. L155 and 177, I would appreciate more clarity on how "evidence for a difference" was determined or defined in this context.

We thank the reviewer for this comment. Considering the Bayesian context of our analyses, we have based this 'evidence for a difference' on the 95% credible interval containing the value 0 or not. In Bayesian statistics, a 95% credible interval is interpreted as 'a 95% probability that the true parameter lies within this interval', given the data and the prior. Hence, it is a direct probability statement about the parameter, incorporating prior beliefs and observed data to update our understanding of the parameter. In case the credible interval does not contain 0, the posterior distribution places most of its mass away from 0, strongly suggesting that there is a difference between the compared groups (e.g. sexual vs. non-sexual transmission). In contrast to frequentist statistics, Bayesian reasoning is not binary (reject vs. don't reject). Instead of p-values, interpretation is in terms of how much probability lies above or below 0, and the strength of that evidence depends on the width of the interval, the posterior distribution shape, and the prior.

2. Figure, The lines in the figure are thin and difficult to distinguish, making it hard to tell which corresponds to sexual or household transmission.

We thank the reviewer for pointing this out and have attempted to make Figure 1 more clear by using different colors for the four types of transmission: household non-sexual (red), household sexual (black), non-household sexual (green), and non-household non-sexual (blue). We believe the Figure is more clear now.

Reviewer #3 (Remarks to the Author):

This is an interesting study on how transmission mode may influence serial intervals and incubation periods in mpox virus transmission. I mainly have questions on the how the data were collected and possible biases.

We thank the reviewer for their interest in our study, and have clarified the data collection process in the revised manuscript based on the reviewer's comments.

1. Line 115: how come that so many contacts were participants of the study? Were they recruited into the study?

Upon study entry, participants were asked about previous contact with mpox-infected individuals, and whether any of them were also participants of the study. The Kamituga study site was the only site offering diagnostic testing for MPXV and therefore, many reported contacts were also participants of the same study.

2. Would it be possible to add a flow diagram or table showing the various percentages of participants in different categories or used for specific analyses? That would be very helpful for the reader.

We thank the reviewer for this suggestion, and have included Supplementary Table S1 to show the number of participants in different categories used for specific analyses.

3. Lines 127-129: What is the rationale for choosing participants who reported only one contact with another participant? Is that not going to introduce a bias into the estimates? Would it be possible to average over several contacts? Also, could these participants have had other contacts outside the study population?

We thank the reviewer for this thoughtful comment. We restricted the non-parametric estimation of the serial interval to participants who reported only one contact with another participant in order to maximize certainty about the direction and identity of transmission pairs. This restriction ensures that the observed serial interval truly reflects a likely transmission pair, without ambiguity introduced by multiple potential sources. We acknowledge that this could introduce selection bias by excluding individuals with multiple contacts, who may differ in important ways (e.g. behavior, exposure risk). However, to mitigate potential bias and to incorporate all available information, we included participants with multiple contacts in our Bayesian model, where probabilistic assignment of infector-infectee pairs is possible. This allowed us to account for uncertainty in transmission chains, and to average over multiple plausible infectors using a fully probabilistic framework. Since results from both approaches are similar, we believe there is no additional bias in our non-parametric estimates besides the biases that are inherent to serial interval estimation in general. An example of such inherent biases is indeed the fact that participants could have had other contacts outside of the study population. Unfortunately, we have no idea about cases hidden in the community as no active case finding was performed.

We have included the rationale behind our choice to only include individuals reporting only one contact in the non-parametric serial interval estimation in the revised manuscript on lines 350-353 and on lines 365-367 as follows:

“For this analysis, we included only participants who reported only one contact, allowing us to assume a likely direct transmission link. This choice minimizes misclassification of infector-infectee pairs, which is a known issue when individuals have multiple possible exposures.”

“For this analysis, we also include participants reporting multiple contacts since the Bayesian framework probabilistically assigns transmission links and accounts for uncertainty in exposure history.”

4. Line 140: You call these pairs transmission pairs, but this relies only on selfreported information about contacts, or was there some way of confirming transmission?

We thank the reviewer for this comment. We indeed refer to these as ‘transmission pairs’ based on self-reported contact information and temporal consistency between symptom onset times, without virological confirmation. In our Bayesian framework we do account for uncertainty in

transmission direction. The non-parametric estimates however were restricted to those individuals with single reported contacts in order to reduce ambiguity, but we agree that without confirmatory data, transmission cannot be definitively established. We have therefore revised the terminology to 'likely transmission pairs' at the appropriate locations throughout the manuscript.

5. Figure 1: It is unclear to me whether the transmission network is based somehow on the assumption that the population is closed, or how it takes into account that contacts can be with persons, who are not participants of the study.

We thank the reviewer for this comment. We did not take into account possible contacts who are not participants of the study. We have added the following to the limitation section in the Discussion on lines 258-261 in the revised manuscript:

"Fourth, there is some inherent bias in the serial interval estimation, as participants could have had other contacts outside of the study population. However, inclusion in the study was near 100% among patients presenting for testing at the only available test site, and we did therefore not account for such unobserved intermediate cases."

6. Line 258: Please give some more background information about this study site. What is the population size?

We have included the following background information about the study sites, including population size, in the revised manuscript on lines 44-52:

"Kamituga is a mining city with a population of around 241 642 and an estimated 20 000 people directly work in the artisanal and small-scale gold mining sector. Mining plays a key role in the livelihood of the local economy, making the city a popular destination that is filled with migrant workers and people searching for economic opportunities. In mid-Jude 2024, Goma, in the North Kivu province of the DRC, experienced its first Clade Ib infections. Goma is a densely populated city of approximately two million inhabitants, comprising more than a dozen refugee camps where nearly 600 000 individuals live in close proximity and precarious conditions due to massive population displacement resulting from ongoing armed conflict crises, making it difficult to take action to control and prevent outbreaks."

7. Lines 266-268: Did the participants name their contacts? Or how could investigators know whether they were study participants?

Participants named their contacts, and subsequently their participant ID was found in the recruitment log available to the investigator on site and participant ID of the contact was then captured in the CRF, pseudonymizing the link. Often, investigators would still know these individuals and the 'infectors' of the pair would have already been admitted in the treatment center.

8. There should be more discussion of any possible biases in the data, for example sampling bias.

We thank the reviewer for this comment and agree that there are indeed some inherent biases in the data used. First, these data were collected during an ongoing outbreak, and consequently right-truncation bias could be present if secondary cases linked to recently infected primary cases have not yet developed symptoms or have not yet been detected. This can result in underestimation of the serial interval because only short intervals are observed

early on. Second, there may be recall bias in self-reported symptom onset. Third, if asymptomatic or mild infections are underdetected, this could lead to selection bias which could overrepresent faster-progressing transmission. However, since the primary aim of our study was comparing the serial interval and incubation period between sexual and non-sexual transmission, and these biases likely affect both types of transmission pairs similarly, the relative difference between groups remains meaningful even if absolute estimates are potentially a bit biased. In addition, common corrections for these biases rely on assumptions about epidemic growth rate, reporting completeness, and incubation period distributions, and as such might introduce more bias rather than remove bias if these assumptions differ between groups. Furthermore, from a public health perspective, the observed serial interval regardless of 'true' infectiousness timing is relevant for outbreak control. There is also a strong possibility that some sexual contacts are misclassified as non-sexual, as suggested by the high presence of anal and/or genital lesions. Since this misclassification is more likely than the other way around (non-sexual misclassified as sexual) the true difference in epidemiological characteristics between sexual and non-sexual transmission could be even larger than what we have observed.

We have added the following to the limitations section in the Discussion, on lines 261-267 in the revised manuscript:

"In addition, serial interval estimation is known to be affected by biases such as right-truncation and recall error. However, these biases are expected to operate similarly across both groups, given similar case investigation methods. Therefore, we focused on observed (uncorrected) intervals to provide a direct, empirical comparison between sexual and non-sexual transmission, rather than applying model-based corrections that require certain assumptions (e.g. epidemic growth rate) that are possibly invalid across differing network structures."

In addition, the following sentence was added on lines 253-255 in the revised manuscript:

"If this is indeed the case, the true difference in epidemiological characteristics between sexual and non-sexual transmission could be even larger than what we have observed."

9. Line 306: Please give some information about what you assume about transmission trees. How are they defined?

We thank the reviewer for this comment and are happy to clarify. In our study, a transmission tree represents a directed acyclic graph where each node corresponds to an mpox case, and edges indicate assumed transmission events from infector to infectee. The structure of the tree captures who infected whom, as well as the timing of those events by means of symptom onset times. We make the following assumptions in defining and inferring transmission trees:

a. Each case has a single infector.
b. Transmission occurs through direct contact between cases in the dataset, i.e. we infer trees under the assumption that all transmission occurred within the observed network of participants.
c. Only transmission links where the observed serial interval (i.e. time between symptom onset of the infector and the infectee) lies within the interval of -5 to 30 days are considered plausible.
d. Reported contacts define the space of possible infectors for each case. When multiple contacts are reported, we treat the identity of the infector as uncertain and model this probabilistically in the Bayesian framework.

We have added the following paragraph about this on lines 389-395 in the revised manuscript:

"We define a transmission tree as a directed set of transmission events among participants, where each node represents an infected individual and each edge represents a likely transmission event from an infector to an infectee. In constructing these trees, we assume that transmission occurs via a direct single exposure between study participants, that the infector's symptom onset precedes the infectee's by a plausible serial interval ranging from -5 to 30 days, and that reported contacts constitute plausible transmission pathways. "

Reviewer #3 (Remarks on code availability):

I did not have time to check the code and reproduce the results. There is a Readme file, which states which scripts are included in the repository.

We have identified a missing piece of code in the R script for the incubation period, and this was added to the repository on 24 July. We have also included an additional R script for the descriptive statistics.