Dear Dr. Narasimhan,

Thank you for your support and continued interest in our paper “Early underdetected dissemination across countries followed by extensive local transmission propelled the 2022 mpox epidemic” This is the resubmission of manuscript CELL-D-23-02872.

Below you will find our point-by-point response letter. Editorial and reviewer comments are shown in black, and our responses in blue.

Yours sincerely,

Miguel Paredes and Trevor Bedford, on behalf of all co-authors

—----------

CELL-D-23-02872

Dear Mr. Paredes,

Thank you for your patience.

I am enclosing the comments that the reviewers have made on your paper. As you will see, the reviewers have commented enthusiastically on the paper, and we would be happy to consider a revised version of the manuscript provided that you satisfy their remaining concerns that are mostly textual and more clarity around the data .

We understand that some revisions take time, but I should mention that we take into account the published literature available on the day we make our final decision. So please make your revisions as soon as possible.

We understand that changes to your circumstances can lead to challenges in completing revisions, especially as various global crises may be causing disruption for you and your colleagues. If that is the case for you and it has an impact on your ability to revise your manuscript, please let us know, and we will be happy to work with you on a plan that works for you to keep your paper moving forward.

In your revision, please include a “Limitations of the Study” section as a subsection of the Discussion. This section is now a standard feature in all Cell papers and it is usually one paragraph long. The goal of this section is to promote clarity and transparency by highlighting any limitations in the interpretation of the study, including limits of the techniques used and/or assumptions made. It can also include mention of additional experiments that would be necessary to definitively prove specific conclusions. These limitations should be specific to the paper.

A “Limitations of the Study” section has been added to the end of the manuscript following the Discussion section.

Before resubmitting, please read our author information concerning stylistic and formatting guidelines, specifically our figure and word count limits. Please note that our article length guidelines have been recently adjusted to reflect that references are no longer factored into the overall word count of the manuscript.

Please be sure to read our editorial policy on authorship annotation, which includes guidelines for co-authorship that have recently been updated. As the complexity and interdisciplinary nature of science grows and evolves, so do the networks of collaborations both within and between labs for published articles. The author list at the start of a manuscript cannot meaningfully convey this increasingly complex information, and we require inclusion of a detailed description of the specific contributions of each author and laboratory in the Author Contributions section of the manuscript.

Figures must be prepared in accordance with the Cell Press Data Processing Policy. For example, if you have digitally eliminated irrelevant or superfluous lanes from a gel or blot image, you must indicate the position of the deletion with a line or a space and explain the manipulation in the figure legend. Details can be found in the Editorial Policies section of our Information for Authors.

The manuscript must also conform to our Supplemental Information Guidelines and our Cell Press methods reporting format, STAR Methods. If you have any questions, please don’t hesitate to contact me.

We’ve consolidated the number of supplementary figures down to 7 in accordance with the Supplemental Information Guidelines and have also updated our methods section to adhere to the STAR Methods format. The Key Resources table is attached separately as requested.

Should your manuscript be accepted for publication in the future, we’ll encourage you to contribute any of these optional features:

- **Figure360**: Create a narrated, animated version of one of your figures that helps the reader zoom in on the most important take-home message in two minutes or less. The video should contain data and panels from only one figure, and include minimal introduction. For further guidelines and examples, please click here.
- **Mendeley Data**: Publish your original, unprocessed data through Mendeley Data. We will link your published paper and the dataset to each other. For more information, please click here.

- **Methods Videos**: We encourage you to make a Methods Video for your paper if you report any methods that are challenging or nuanced, or if you have an experimental setup that is hard to describe. These videos are short (≤1 min) and are intended to improve reproducibility and transparency. For examples and guidelines, go here.

We appreciate that there is a lot of information in the instructions, so if you have any questions on any of these policies, don't hesitate to contact our Editorial Operations Associate, Michelle Kubilis (cellms@cell.com).

To resubmit, you can log in into your EM account here. I look forward to hearing from you and to reading the revised manuscript.

Complement your primary research article by publishing a step-by-step procedure with STAR Protocols, an open-access peer-reviewed journal from Cell Press. STAR Protocols aims to make the daily work of the scientific researcher easier by providing complete, authoritative, and consistent instructions on how to conduct experiments. The primary criteria for publication in STAR Protocols is usability and reproducibility. You can check out their most recent protocols here. If you have any questions, please email starprotocols@cell.com.

Best wishes,

Sri Devi Narasimhan, Ph.D.

Deputy Editor, Cell

Reviewers' Comments:

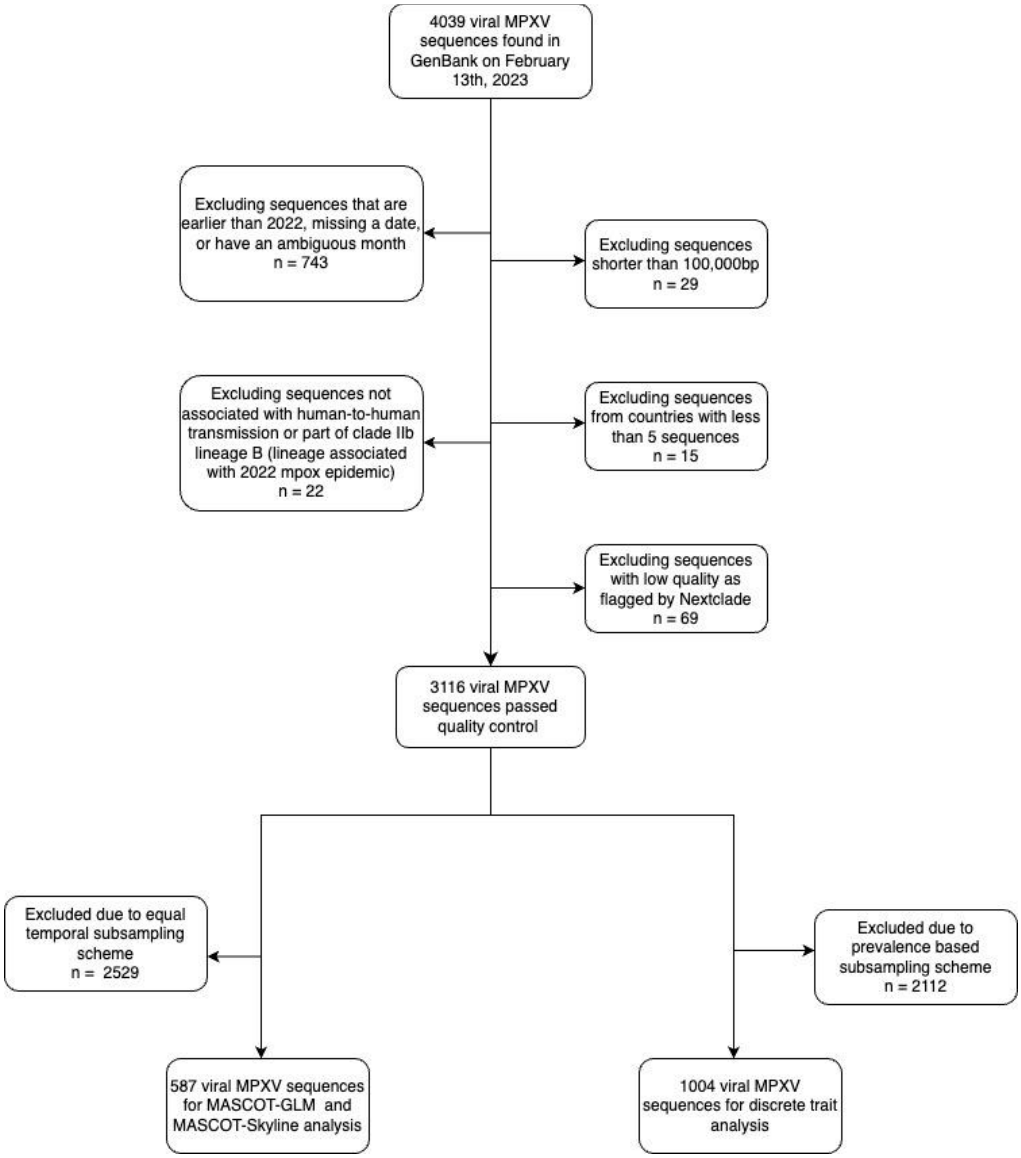
Reviewer #1: General comments:
This is a Research Article by Paredes et al, using publicly available Mpox genomes from across the globe between March and December 2022 to perform scientific and policy relevant phylodynamic and phylogeographic analyses to retrospectively understand key epidemiologic questions. The key findings include: under-detection of Mpox cases early in the pandemic, majority of transmission occurred due to local transmission rather than importation of cases, and cases declined prior to large scale vaccination in the US. Overall, this is a compelling and well conducted study and an important contribution to the field, especially as a case example of how these methods can be applied during an outbreak. Thank you to the authors for their important study. I hope my additional minor comments can be constructive.

We thank the reviewer not only for their kind words but also for their constructive comments that improved the manuscript!

Major comments:
--Sample selection and bias: The study includes genomes from ~1% of cases from the outbreak (1004 sequences out of 88,549 cases), and given case under-ascertainment, this fraction is likely far smaller. Furthermore, there is bias in which viral isolates are sent for sequencing (i.e., this is not a random sample); there are likely changes over time and selection bias at a finer spatial scale (perhaps even sub-state level in the US) based on capacity. This is not meant as a critique, as the use of publicly available genomic data for this analysis is the only reasonable approach. The authors also do a commendable job to evaluate this limitation and the study overall is very well done. However, two additions would strengthen the presentation of data to address this: (1) Addition of a new Figure with a flow chart of the sample size of genomes used in the main analysis with inclusion/exclusion criteria, starting with the number of known cases (similar to a PRISMA flow chart). E.g., How many genomes in GenBank? How much missing metadata affect sample size?;

We agree that nonrandom sampling, the changing availability of genomic sequencing, and unequal sampling across the regions study affect the probability that a case shows up as a sequence in our dataset through the period studied and thus might bias our results. We attempted to account for this variation by weighting the subsampling for our phylogeographic (DTA) analysis according to confirmed case counts, and by oversampling undersampled regions (and downsampling overrepresented regions) in our MASCOT-GLM analysis, as well as by adding in case counts as an empirical predictor in the model in an effort to account for this variation. The approximate structure coalescent approaches have also been found to be less susceptible to sampling bias, especially when augmented with relevant predictors.

We also agree that a flow chart of the sample size of the genomes used in our two main analyses with inclusion and exclusion criteria would strengthen the manuscript. As such, we have included the following diagram as part of Supplementary Figure 1 and have referenced it throughout relevant sections in the text.


4039 viral MPXV sequences found in GenBank on February 13th, 2023
Excluding sequences that are earlier than 2022, missing a date, or have an ambiguous month
n = 743
Excluding sequences shorter than 100,000bp
n = 29
Excluding sequences not associated with human-to-human transmission or part of clade IIb lineage B (lineage associated with 2022 mpox epidemic)
n = 22
Excluding sequences from countries with less than 5 sequences
n = 15
Excluding sequences with low quality as flagged by Nextclade
n = 69
3116 viral MPXV sequences passed quality control
Excluded due to equal temporal subsampling scheme
n = 2529
Excluded due to prevalence based subsampling scheme
n = 2112
587 viral MPXV sequences for MASCOT-GLM and MASCOT-Skyline analysis
1004 viral MPXV sequences for discrete trait analysis

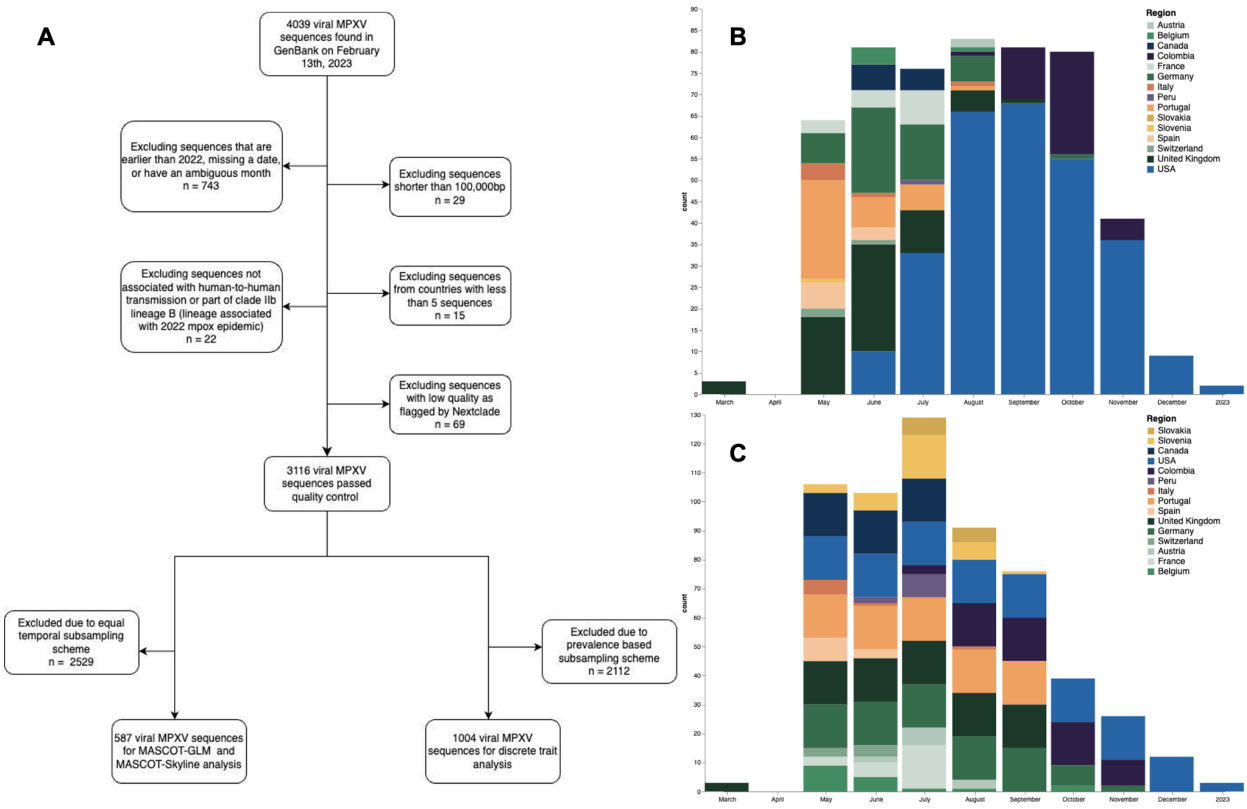


Figure S1: Subsampling for phylogeographic and phylodynamic inference, related to Figure 1. (A) Flow diagram displaying the inclusion and exclusion criteria for the final two analytic samples (B) Temporal Distribution of 1004 genomes used for phylogeographic analysis. Genomes were subsampled using confirmed case counts as weights. (C) Temporal distribution of 587 genomes used for MASCOT-GLM analysis. Subsampling was done to promote an equal number of samples from each deme for each month in order to oversample underrepresented countries.

(2) Expansion of Limitations paragraph in the Discussion to guide reader how biases (probability of being sequenced) may affect the direction of the key study findings. The current discussion is helpful but vague on how these biases would affect the results; more guidance on which direction this bias will affect the results (if known) would be helpful. As one example, could the authors elaborate on how sampling bias affects the analysis on transmission heterogeneity? This could also be done in the Appendix if space is limited.

We agree that the limitations paragraph could be expanded to discuss the impact of sampling bias in our study, as such we have included the following additions to the "Limitations of the study" section:

> *Additionally, the changing availability of genomic sequencing, as well as unequal sampling across the regions study affect the probability that a case shows up as a sequence in our dataset through the period studied. If viruses migrate frequently between*

our study countries and countries that lack genomic sampling, the lack of samples that might interdigitate with samples from the study country may affect our ability to distinguish separate introductions. Despite this potential bias, the 2022 mpox epidemic mainly affected Europe and the Americas, which are regions that are well represented in our study, limiting the effect of this bias.

As well as:

In the analysis of transmission heterogeneity, we explicitly accounted for the fraction of cases sequenced and explored several assumptions regarding the proportion of infections detected by the surveillance system. This was done assuming that all infections had the same probability of being detected as cases and sequenced. Active surveillance targeting larger clusters could lead to underestimating the extent of transmission heterogeneity

-US vaccine campaign analysis: While I believe the authors' conclusion is valid, I would suggest consideration of a few minor points to improve this analysis. (1) The population definition used to estimate vaccine coverage (MSM with PrEP, MSM with HIV) and to calculate R_t (entire population in North America) is different. The assumption when interpreting Figure 6 is perfect mixing, and compatible population definitions. However, the degree of transmission heterogeneity suggests this may be more complicated. If the authors are able to address this with revising the samples included in R_t analysis (or revising vaccine definition), or simply adding this as a limitation, all options are fine.

We agree that the incongruent population definitions and the violation of the assumption of random mixing are limitations for our analysis. While we attempted to reconcile the populations, we were unable to find publicly-available vaccination data for Canada or publicly available metadata associated with MPXV sequences that identified membership of a high risk group. However, mpox in the US and Canada spread predominantly among high-risk MSM populations, with cis-gender women accounting for only about 2.6% of cases (1). Given that there is no reason to expect that the sequenced cases would be significantly different from the demographic distribution of cases, we believe that the majority of the North American sequences in our study were derived from a similar population as we used to estimate vaccine coverage. We highlight these limitations in two different paragraphs within the Discussion section:

The incongruent population definitions (high-risk population in the US for vaccine-derived immunity and a single Rt estimate for the US and Canada combined) could conceivably bias our conclusions regarding the relationship between Rt and vaccine coverage. Despite the lack of publicly available vaccination information for the whole of Canada, vaccination data for Montreal (46) and Ontario (47) show that pre-exposure vaccination began at a similar time, if not later, than the vaccination efforts in the US (immunization in Montreal began on May 27, 2022 and in Ontario on June 9,

2022 compared to May 22, 2022 in the US (4)). We believe that the similar, if not later, start of vaccination campaigns in Canada biases our results in a conservative and limited fashion compared to what would be expected if the Canadian vaccination efforts began earlier than those in the US. In addition, the Canadian MSM population is estimated to be 10% of the US population (48,49), further suggesting that vaccination in Canada should have a limited role in reducing mpox Rt in North America. More broadly, our estimates of Rt and vaccine-derived immunity aggregate across large spatial regions. Further spatially resolved analyses could provide additional information about the relationship between Rt and vaccine coverage.

—

Bayesian coalescent models assume random sampling of infected individuals, meaning that targeted sampling of superspreader events, or via contact tracing, could bias our phylodynamic estimations. We attempt to quantify the extent of transmission heterogeneity via our estimates of overdispersion (Figure 7). In the analysis of transmission heterogeneity, we explicitly accounted for the fraction of cases sequenced and explored several assumptions regarding the proportion of infections detected by the surveillance system. This was done assuming that all infections had the same probability of being detected as cases and sequenced. Active surveillance targeting larger clusters could lead to underestimating the extent of transmission heterogeneity (23,58).

1. *Blackburn D, Roth NM, Gold JAW, Pao LZ, Olansky E, Torrone EA, et al. Epidemiologic and Clinical Features of Mpox in Transgender and Gender-Diverse Adults — United States, May–November 2022. MMWR Morb Mortal Wkly Rep. 2022 Dec 30;71(5152):1605–9.*

(2) Related point on the spatial scale of both vaccination and transmission. The high level of spatial aggregation for this analysis (entire US for vaccine and North America for R_t) preclude detection on more meaningful changes occurring at finer spatial scales, e.g. city level where vaccine coverage may be higher than the national average. This could be a scenario analysis looking at 2-3 large US cities, or adding this as a limitation.

We agree that the high level of spatial aggregation in this analysis obscures fine scale resolution of mpox transmission and vaccination coverage. However, given the computational complexity of increasing the number of demes in our Bayesian phylodynamic analyses, we chose to focus on large global regions rather than sub-country regions to understand global transmission patterns.

Given that the focus of the manuscript is to understand global and regional trends, we believe that doing a sub-country analysis would be out of scope of the manuscript as well as highly restricted due to the scarcity of publicly available fine-scale mpox vaccination data. We highlight the limitation of the high level of spatial aggregation and the potential for further research via inclusion of the following statement in the discussion paragraph that highlights the vaccination coverage results:

> *More broadly, our estimates of Rt and vaccine-derived immunity aggregate across large spatial regions. Further spatially resolved analyses could provide additional information about the relationship between Rt and vaccine coverage.*

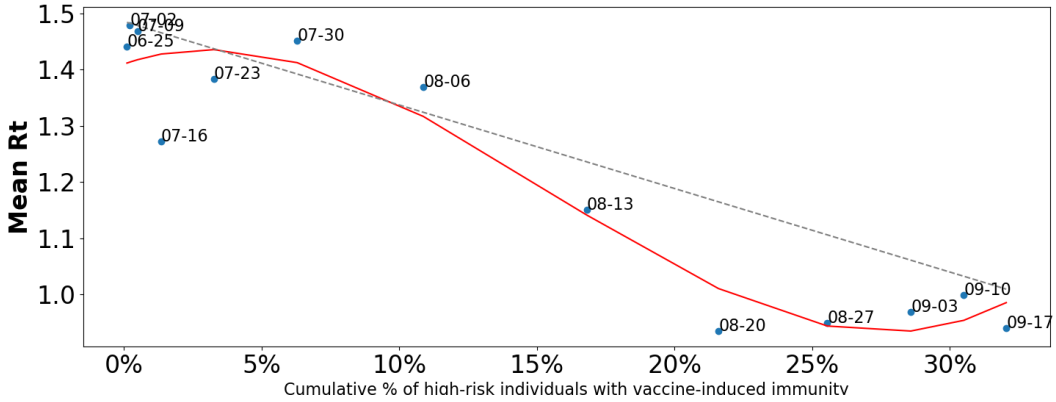
(3) Figure 6: Consider adding a second line without the 2-week lag in immunity as a sensitivity analysis;

We agree that this sensitivity analysis would strengthen our conclusions. Below you'll find the same analysis presented in main figure 6 but repeated with no lag for immunity since the date of vaccination as well as with only a one week lag since the date of vaccination. We can still see a clear trend of empirical *Rt* in red appearing below the dashed line expectation from vaccine-derived immunity. In the scenario of immediate immunity after vaccination, we still observe that measured *Rt* drops below 1 before ~22% of the high-risk US population acquire vaccine derived immunity while we expect that it would require ~33% of the high-risk population to have vaccine derived immunity to reach an *Rt* less than 1 (dashed line meets *Rt* of 1 at this point). These results lend support to the robustness of our conclusion and were included as part of new Figure S6C-D.

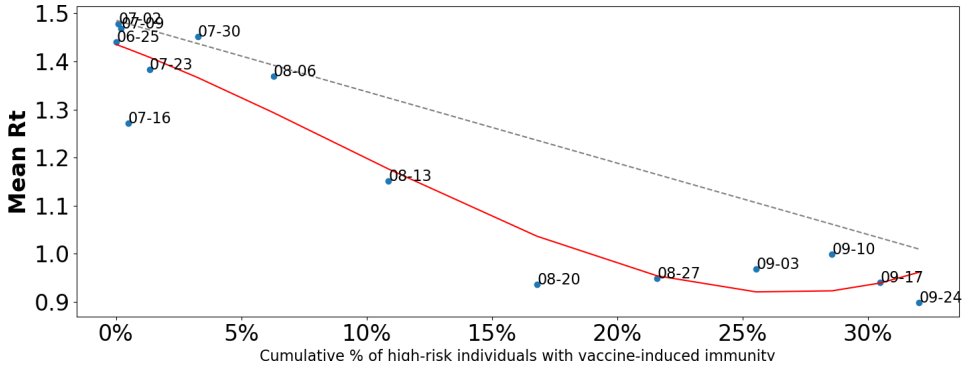
Reference to the supplementary figures are as follows:

> *The decay of Rt in North America before a substantial percentage of high-risk individuals developed vaccine-related immunity remains clear even when assuming no lag or a one week lag after vaccination for the development of immunity (Figure S6C-D)*

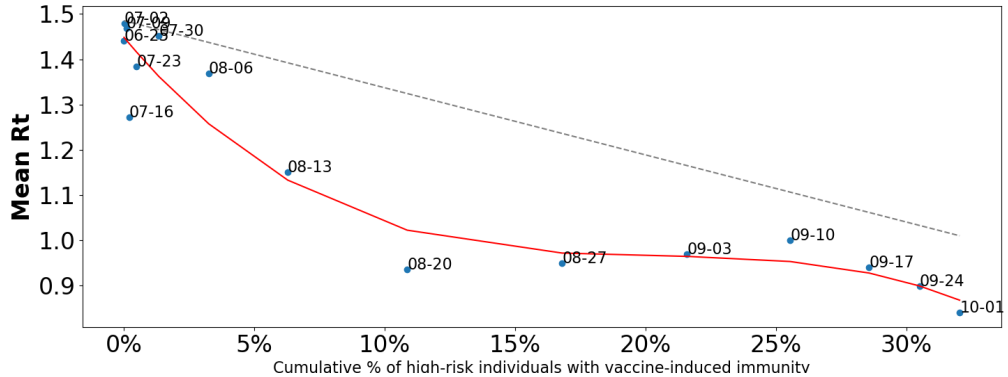
No lag for immunity:



One week lag for immunity:



Original (two week lag since immunization):



New figure S6:

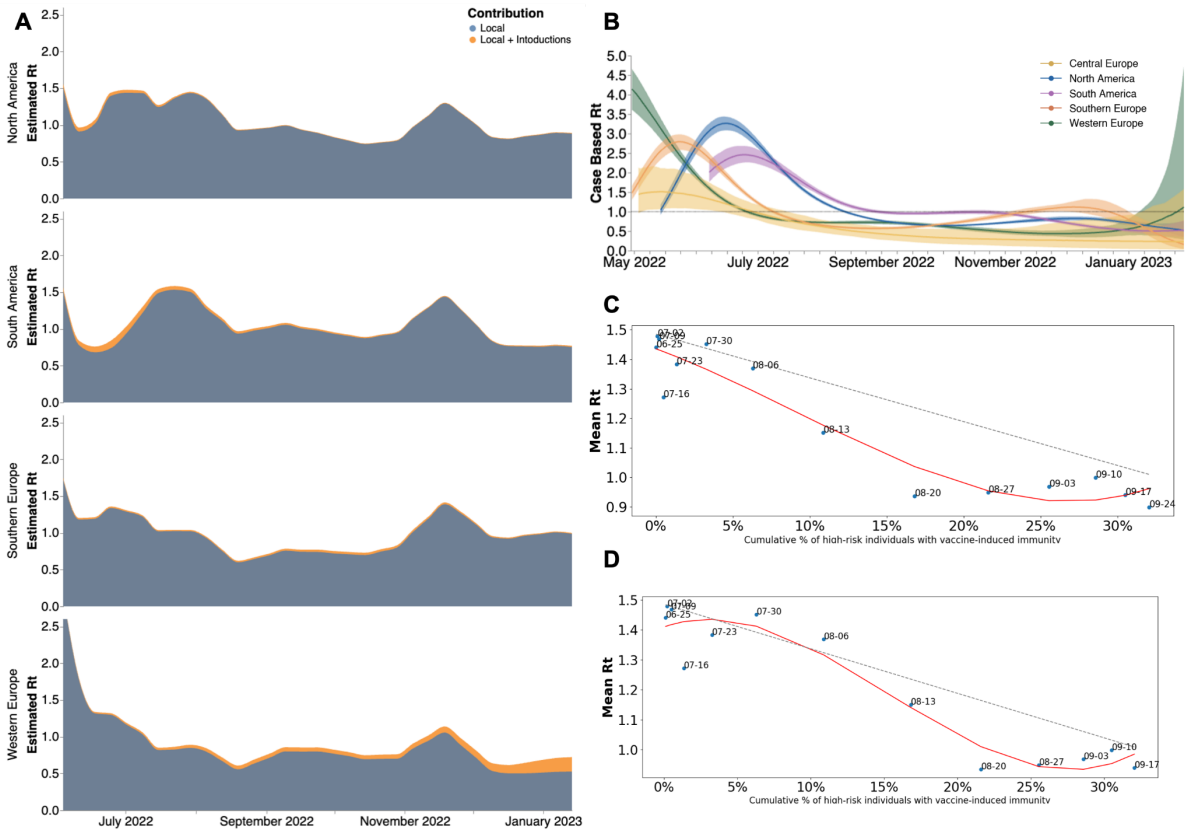


Figure S6: Estimates of time-varying reproductive number (Rt), related to figures 5 and 6. (A) Estimates of Rt from April 2022 through December 2023 via MASCOT-GLM for four global regions separated by source of contribution. Blue denotes local Rt without the influence of outside viral introductions while orange shows the added contribution of introductions. Central Europe was removed due to limited data on introductions. **(B)** Estimates of time-varying reproductive number (Rt) in five global regions Estimates of Rt from April 2022 through December 2022 using renewal model framework from case counts only. The inner area denotes the 50% HPD interval and the outer area denotes the 95% HPD interval. Dashed line highlights an Rt value of 1 above which denotes an exponentially growing viral epidemic. (**C-D**) Scatter plot comparing mean *Rt* calculated via MASCOT-GLM for North America vs cumulative percentage of high risk individuals with vaccine-induced immunity in the United States with a one week lag to account account for the development of immunity **(C)** and no lag following date of vaccination **(D).** Red line indicates the best fit spline for scattered points. Dashed gray line indicates expected linear decrease in *Rt* with increasing vaccine-immunity assuming SIR dynamics. Over each point are the dates that correspond to the mean *Rt* and percent of immunity at that moment.

(4) Is natural infection accounted for here?

We only account for vaccine-derived immunity but believe that natural infection would have had a limited impact on our estimates based on the estimated proportion of high-risk MSM in the US that had a reported case of mpox. We expand on this limitation in the discussion section of the manuscript:

In the present analysis, we had only accounted for vaccine-derived immunity. However, we can attempt to account for immunity derived from natural infection by comparing mpox cases to the size of the at-risk population. Doing so, we find that less than 2% of the high risk MSM population in the US had reported cases of mpox as of Nov 22, 2023. Converting this crude cumulative incidence into an estimate of the total proportion of the population infected requires knowing the reporting rate of mpox infections. If we assume complete reporting then we expect just 2% cumulative incidence, which should have negligible impact on lowering epidemic Rt. However, if the reporting rate was 10% then we expect approximately 20% cumulative incidence, which starts to have an impact on Rt. While we were unable to find a precise estimate of reporting rate in the US, prior studies in Portugal (44) and in North Carolina (45) estimate the rate of detection to be 62% (95% CI, 43%-83%) and 66% (95% CI, 44%-91%). If we assume that the US reporting rate falls on the lower bound of those estimates, we expect 4.7% cumulative incidence. Thus, in general we expect that natural immunity will have played a minor role on reducing epidemic Rt compared to behavioral modification and vaccine-derived immunity.

(5) Is the date of mpox case in this dataset the date of sample collection, symptom onset, date of positive test, sequencing, etc? I mostly ask since the R_t drop and vaccination are quite close together, so want to ensure the dates are correct and consistent given use of secondary data.

The dates used in the sample for the MASCOT-GLM analysis from which the *Rt* estimates are derived refer to the date of sample collection of an mpox infection that resulted in an included sequenced genome. Our estimates of *Rt* are based on our structured coalescent estimates of effective population size, which require the date of specimen collection for each sequence. Estimates of effective population size allow us to estimate the size of the viral population through time and thus are calibrated based on the moment of viral collection. This has clarified in the Methods (such as in *Sequences with ambiguous date of collection in the month column, with a sample collection earlier than January 2022, and flagged as being low quality by Nextclade*) and a clarification about the dates in figure 6B was also included in the legend (*Over each point are the dates that correspond to the mean Rt and percent of immunity at that moment*.)

For the development of vaccine-derived immunity, we focused on the date of first mpox vaccination. We describe the method as follows in the STAR Methods:

> *In order to account for the development of immunity, we followed the CDC method of assuming the development of immunity took two weeks following vaccination and thus only considered individuals as "having vaccine-induced immunity" after reaching two weeks from the date of first vaccination.*

--Comparison of R_t estimates between epidemiologic data and genomic data. The authors note that the standard R_t estimate reliant on epi data alone (Figure S9) yields poor estimates of R_t, especially early on in the pandemic, compared to phylodynamic estimation (Figure 5). This is a really interesting finding and highlights a strong use case for genomic surveillance, especially early in a pandemic to derive this important public health parameter. However, does this discrepancy persistent once case ascertainment is higher and stable over time? The analyses suggest no based on my read, but if the authors can comment here, it would be a useful finding to clarify. Would then also be punchier in the Discussion to highlight the use of these methods, but be open that later in outbreak, the difference is not as apparent (if this is the case)

We thank the reviewer for their insights. Yes, after initial period of under-detection we expect that case-based Rt will be accurate. We provide more explanation and clarification on this topic in the discussion as follows (additions highlighted in **bold**):

> *By contrast to Rt estimates obtained solely from case counts (Supp. Fig 9 or (9)), we obtained estimates that are smaller in magnitude relying on phylodynamic models informed by prevalence estimates and monthly predictors that account for changes in case detection.* ***Case-based Rt calculations will be overestimated if case detection is increasing as the estimates capture both the true rise in infections and the rise in detection of infections. For mpox, case-detection stabilizes in August 2022 (Fig. 3D), after which time, we expect case-based Rt estimates to be more accurate. This suggests that our approach of integrating multiple data sources would provide a more accurate estimation of mpox transmissibility and of the impact of interventions, especially in the beginning stages of an epidemic where accurate knowledge of Rt can have high impact in informing public health action****.*

Minor comments:
--Abstract: 1) Suggest tempering the causal language ("effect"), given the methods used in the analysis. 2) Consider changing "population-level intervention" to either "vaccination" (which you evaluated) or "population-level changes" (to be broader and allow for both behavior change and vaccination). 3) Conclusion is a bit bland. Revise to focus on role of genomics during outbreaks or the key policy implications, I defer to authors here.

Thank you for these suggestions. We removed “effect” and focused on associations. Additionally, we changed “population-level intervention” with “population-level changes” and strengthened the conclusion.

--Figure 3: Why are there only 587 sequences instead of the 1004?

In addition to the new supplementary figure S1 that shows the flowchart of sample size of genomes, we clarify the difference in the number of sequences between the two approaches in the results section:

> *Despite the improved computational efficiency of MASCOT over standard structured coalescent approaches (10), parameter inference under MASCOT-GLM is still computationally demanding compared to discrete trait analysis. For reference, the runtime for our main DTA analysis with 1004 sequences was about 12.26 hours/million states while for MASCOT-GLM with only 587 sequences it was about 16.45 hours/million states. Requiring* $5 * 10^7$ *MCMC steps to promote convergence, these runtimes translate to 25.5 days of computational demand for our main DTA analysis and 34.3 days for our main MASCOT-GLM analysis. As such, we reduced the number of sequences to 587 for MASCOT-GLM to allow for inference within actionable timescales.*

--What explained difference in MRCA estimates (March vs Nov-Dec 2021)? Has outside literature investigated this? That's quite a difference. It's only mentioned briefly in the Discussion.

We thank the reviewer for their suggestion. We have expanded our discussion on the observed differences in TMRCA in the Limitations section as follows:

We see a discrepancy in the TMRCA between various models (Table S2) and find our estimates to be highly dependent on the tree prior and thus should be interpreted with caution. Inference of TMRCA is dependent on the estimate of effective population size in early 2022. Different tree priors assume different parametric forms of effective population size and so differ in TMRCA estimates. The rapid exponential growth observed in early 2022 suggests that effective population size should be low in January-March 2022. This information is used by the DTA skyline and skygrid models, as well as the MASCOT-skyline model, resulting in TMRCA estimates close to the earliest March sequences. Consistently, the MASCOT-GLM model estimates the coefficient of the monthly predictor for April 2022 and earlier at -1.09 (95% HPD: -1.89 – 0.00, Fig. 3D), again supporting a small effective population size in this time period. We suggest a conservative interpretation of these results supporting a TMRCA between September 2021 and March 2022.

--Sub-sampling: Is this sub-sampling done once (with set seeds) or is it repeated >N times to determine robustness of the random sampling here?

Subsampling was done with set random seeds in order to promote reproducibility of the results. These random seeds can be found in the github repo such as in https://github.com/blab/mpox-dynamics/blob/main/monkeypox-build/config/config_hmpxv1.yaml#L24 for example, and has been clarified in the methods: *For this analysis, we considered each global region as a discrete location and employed subsampling weighted by mpox case counts for each region using a random seed, resulting in a final subset of 1004 sequences (distribution across countries and regions shown in Supplementary Table 1).*

Despite the computational burden of our Bayesian phylodynamic analyses limiting our ability to analyze a large number of different subsamples, the sensitivity analyses conducted in Supplementary Figures 3 and 4 were conducted with both different random seeds and different subsampling schemes and resulted in highly similar results, lending stronger support to the robustness of our subsampling than would be achieved via the same subsampling scheme with different set seeds.

--Site masking: Why did the authors mask 90% instead of 99.9% if only 0.01% were polymorphic?

The manuscript incorrectly reported the percentage of polymorphic sites as less than 0.01% while the true percentage is less than 1% (~0.8%). The error arose from a confusion when converting decimals to percentages. This has been corrected in the Methods section under the subheading *Site Masking*. We thank the reviewer for highlighting this. We chose to mask only 90% instead of 99% as we found the improvements in computational efficiency to be comparable and allows for easy adjustment of the clock rate by a magnitude of 10.

--Figure 3D: Which predictor coefficient is this for the model? Some additional clarification in figure legend would be helpful.

All of the predictors in Figure 3D were included in the analytic model. We have clarified this in the figure legend as follows: *All of the predictors displayed on the x-axis were included in the main analytic model.*

--R_t estimate (page 21): What is implication of assuming single variant for this approach?

The phrase “assuming a single variant” in the Methods section for the Rt estimates refers to the lack of selection within a single population. We agree that it is confusing and we have removed it from the manuscript as it was found to be superfluous.

--Is there new literature to compare the evolutionary rate found in this study?
Prior literature has found a higher than expected evolutionary rate for both the B.1 and A lineages of clade IIb. To our knowledge, these publications only represent the increased rate of lineage B.1 by calculating the mean number of SNPs between MPXV from 2022 vs 2018-19, such as in Isidro et al. (2022), while our estimates are inferred directly from our Bayesian phylodynamic analyses. Other publications have reported the evolutionary rate of lineage A, which is also part of clade IIb like lineage B.1 studied in our manuscript. Gigante et al. (2022), find the evolutionary rate of lineage A to be ($7.2 \times 10^{-6} \pm 8.9 \times 10^{-7}$ SD) while O'Toole et al. (2023) report the evolutionary rate of the entire IIb clade but divided by APOBEC3 (1.2×10^{-4}) and non-APOBEC3 (4.2×10^{-6}) partitions, making a direct comparison with our results difficult although our estimates of fall within that range. We have cited these sources throughout our manuscript.

Isidro J, Borges V, Pinto M, Sobral D, Santos JD, Nunes A, et al. Phylogenomic characterization and signs of microevolution in the 2022 multi-country outbreak of monkeypox virus. Nat Med. 2022 Aug;28(8):1569–72.

Gigante CM, Korber B, Seabolt MH, Wilkins K, Davidson W, Rao AK, et al. Multiple lineages of monkeypox virus detected in the United States, 2021–2022. Science. 2022 Nov 4;378(6619):560–5.

O'Toole Á, Neher RA, Ndodo N, Borges V, Gannon B, Gomes JP, et al. APOBEC3 deaminase editing in mpox virus as evidence for sustained human transmission since at least 2016. Science. 2023 Nov 3;382(6670):595–600.

Thank you to the authors for the opportunity to review this important work.

Reviewer #2: This study analyzes a large, publicly available dataset of MPXV genomes from the 2022 worldwide outbreak. The analyses rely on a mixture of phylogeographic, phylodynamic, and generalized linear model regression techniques. Results include estimates of TMRCA, effective population size, Rt, number of introductions, and transmission overdispersion, and they point to the importance of early, undetected transmission in the outbreak, the relatively small role of worldwide viral migration after the earliest introductions, and the likely larger role of behavioral modification than of vaccination in the initial decline of the outbreak.

The authors display expertise in a number of powerful analytical tools, many of which they

have helped develop. They make commendable efforts to test the robustness of their conclusions at multiple points by, for example, testing the inference of the geographic source of samples by masking the known origin of 10% of samples, by repeating analyses with different approaches, and by assuming a range of undersampling for their dataset. If there are substantial problems with their approach, they lie beyond my competence to identify. The manuscript is also generally quite well written.

We are grateful to the reviewer for their careful consideration of our paper, their constructive suggestions, and their support!

I have no real concerns about the manuscript, just a suggestion for making it more broadly interesting. The mpox outbreak was unusually well sequenced for a large global epidemic and the authors have taken multiple approaches to modeling its dynamics. This would seem to be a good opportunity to discuss their concordance quantitatively, especially for a non-specialist audience. For example, the TMRCA estimates varied substantially between methods, while agreement for various parameters was more mixed between estimates with and without non-phylogenetic predictors. I (at least) would find it interesting to see a summary of such comparisons, along with any thoughts the authors might have about their meaning. How seriously should HPD intervals be taken for TMRCAs? Just a thought - the paper is fine as it is.

We thank the reviewer for their suggestion. We have updated Table S2 with a wider array of summary statistics (TMRCA, Migration Rate, and Clock Rate) as well as with a larger number of varying tree priors with and without predictors as seen below to allow for easier comparison of model choice and of discussion on how tree priors after our inferences.

Table S2. Comparison of time to most recent common ancestor (TMRCA), migration rate (migration events per year), and clock rate (substitutions per site per year) by method, related to Figures 2 and 3. First five rows denote the comparison of key summary statistics from main and alternative models used (with varying tree priors and inclusion of empirical predictors) which include three sequences from March 2022 which were found retrospectively in the UK. The last two rows represent the main analyses but without the three retrospective march 2022 samples.

Analysis	Tree Prior	TMRCA		Migration rate (events per year)		Clock rate	
		Mean	95% HPD	Mean	95% HPD	Mean	95% HPD
With three retrospectively-collected March 2022 sequences from the UK							
MASCOT-GLM	Approximate	2021-12-03	(2021-09-21	1.77	(1.45	6.27E-05	(5.62e-5

	Structured Coalescent		to 2022-02-01)		-2.13)		– 6.96e-5)
MASCOT-Skyline	Approximate Structured Coalescent	2022-01-29	(2021-12-07 to 2022-03-12)	1.83	(1.38-3.00)	5.71E-05	(5.06e-5 – 6.39e-5)
DTA	SkyGrid	2022-03-24	(2022-03-09 to 2022-03-27)	0.72	(0.57-0.89)	8.41E-05	(7.71e-5 –9.10e-5)
DTA	Skyline	2022-03-18	(2022-02-25 to 2022-03-27)	1.34	(1.03-1.66)	5.64E-05	(5.00e-5 – 6.39e-5)
DTA	Constant	2022-01-23	(2021-12-21 to 2022-02-24)	0.81	(0.68-0.89)	9.44E-05	(8.64e-5 –1.05e-4)
Without March 2022 sequences from the UK							
MASCOT-GLM	Approximate Structured Coalescent	2021-11-29	(2021-09-19 to 2022-01-28)				
DTA	SkyGrid	2022-03-30	(2022-03-05 to 2022-04-19)				

We reference table S2 in the results "*Alternative phylodynamic models place the TMRCA in November or December 2021, but show similar estimates of clock rate (Table S2).*"

Additionally, we have expanded our discussion on the observed differences in TMRCA in the Limitations section as follows:

We see a discrepancy in the TMRCA between various models (Table S2) and find our estimates to be highly dependent on the tree prior and thus should be interpreted with caution. Inference of TMRCA is dependent on the estimate of effective population size in early 2022. Different tree priors assume different parametric forms of effective population size and so differ in TMRCA estimates. The rapid exponential growth observed in early 2022 suggests that effective population size should be low in January-March 2022. This information is used by the DTA skyline and skygrid models, as well as the MASCOT-skyline model, resulting in TMRCA estimates close to the earliest March sequences. Consistently, the MASCOT-GLM model

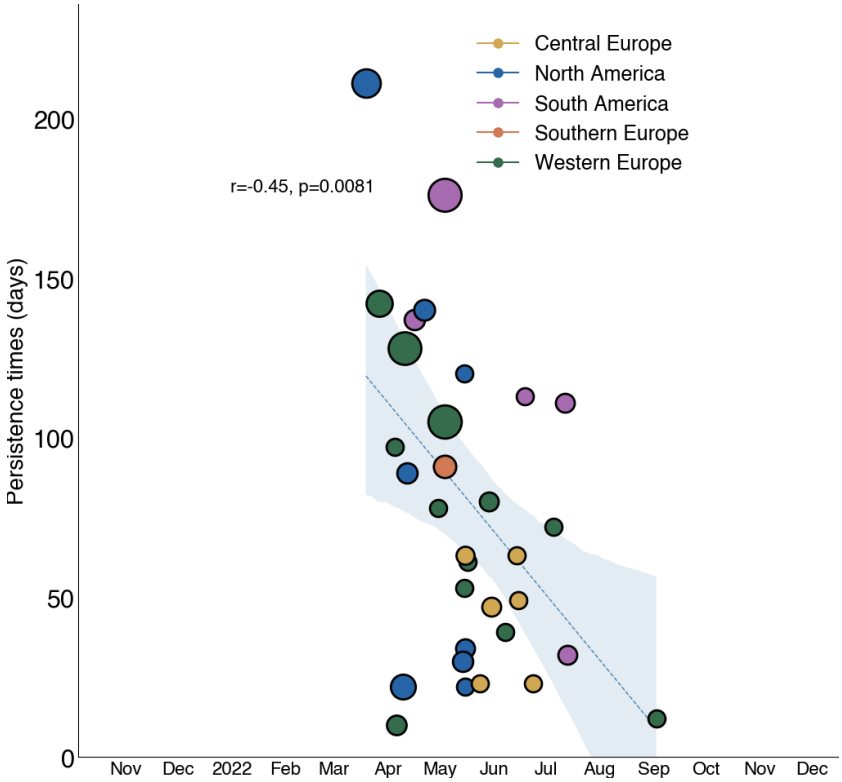
estimates the coefficient of the monthly predictor for April 2022 and earlier at -1.09 (95% HPD: -1.89 – 0.00, Fig. 3D), again supporting a small effective population size in this time period. We suggest a conservative interpretation of these results supporting a TMRCA between September 2021 and March 2022.

How reliable were the phylodynamic and phylogeographic inferences without the case number and air travel data, for example?

We tackled this issue in multiple ways. For the phylodynamic analyses where MASCOT-GLM was informed by the predictors, we repeated the analysis using only genomic sequence information via MASCOT-Skyline, which uses the same approximate structured coalescent approach as MASCOT-GLM but without the influence of predictors. These results are shown in Figure S5 as show similar patterns of effective population size but with more uncertainty, supporting our choice of using MASCOT-GLM for our main analysis. Our phylogeographic inferences do not include predictors in the model but the subsampling is weighted by case counts. To test the robustness of the subsampling, we repeated the phylogeographic analysis using two different subsampling schemes independent of case counts and we found highly similar results. These results are presented in Figure S3.

One small question: To what extent is the trend seen in Fig. 4A driven by the two early introductions? Is there still a significant trend without them?

To answer the reviewer's question, we repeated the analysis seen in Fig.4A but without the first two early introductions as seen below. We find that the trend remains significant, showing a negative correlation between timing of introduction and persistence time of the resulting transmission chain. The correlation is slightly weaker ($r = –0.45$ vs $r = –0.69$ in the original analysis) suggesting that these first two introduction events played an impactful role in promoting epidemic spread, most likely due to high susceptibility and a high degree of underdetection from public health surveillance. This sensitivity analysis has been included as part of Figure S4C



We also included the following description in the results section in order to reference the above sensitivity analysis:

> *The clear negative correlation between time of introduction and persistence of downstream transmission chains remains even without the influence of the two large transmission chains following the first two inferred introductions (Figure S4C)*

New Figure S4:

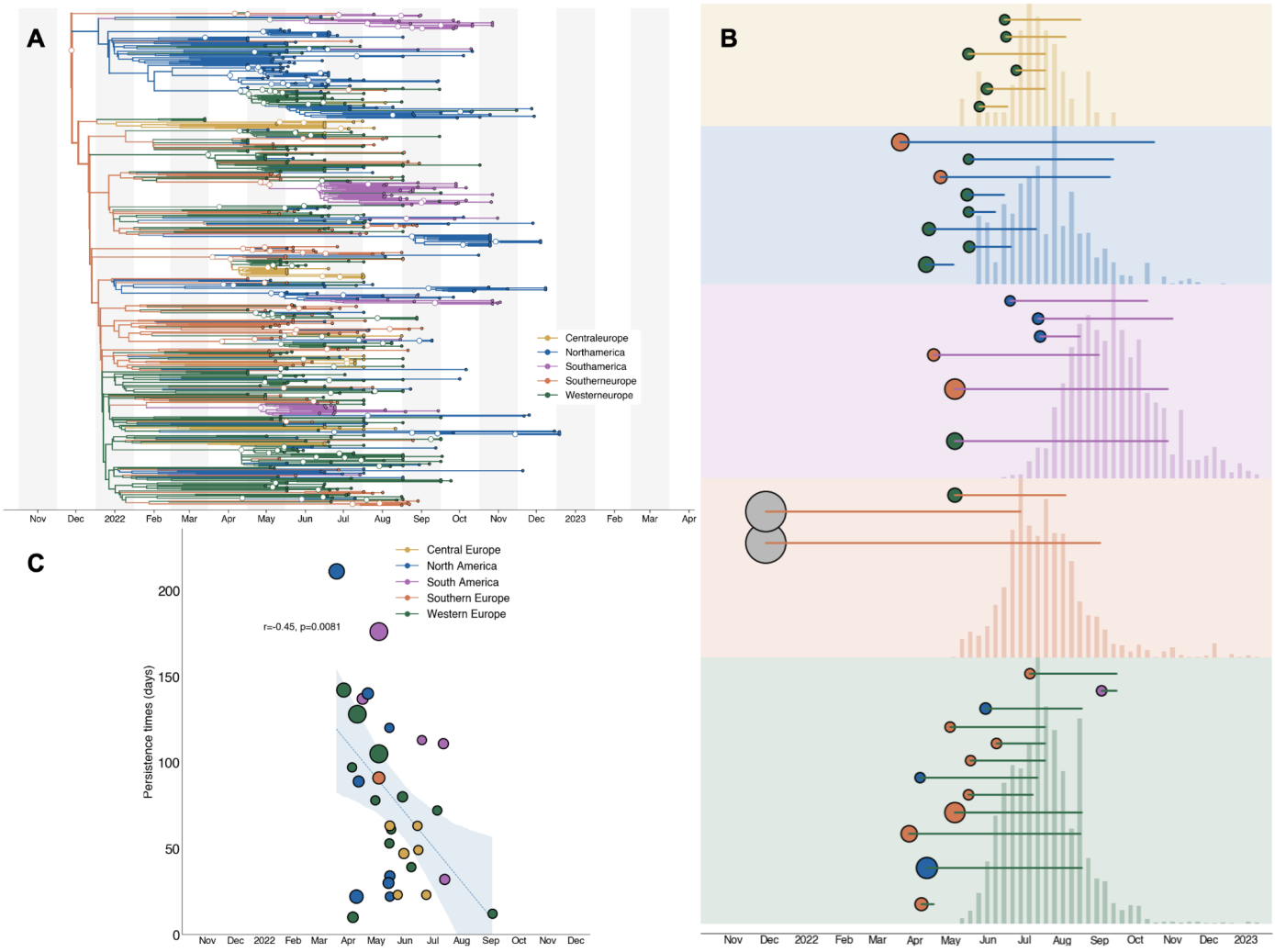


Figure S4: Analysis of introductions inferred via MASCOT-GLM, related to Figure 3 and 4 (A) The maximum clade credibility tree summary of the Bayesian inference conducted using MASCOT-GLM on 587 sequences. Colors correspond to the regions in the legend. Ancestral nodes with greater than 50% posterior support are highlighted with a white circle overlaid. **(B)** Exploded subtrees for each region with only the introductions with greater than 50% posterior support showing that early underdetected introductions lead to longer transmission chains. Color at introduction origin represents inferred source region and size of the circle at the origin is proportional to the number of downstream tips. Length of line coming out of each introduction origin represents the length of the transmission chain. Case counts are overlaid for each region .**(C)** Relationship between estimated date of introduction and persistence time with the first two large introductions removed. Each circle represents a single viral introduction with greater than 50% posterior support into the region denoted by the color (i.e. a green point represents an introduction into Western Europe). The size of each point is proportional to the size of the outbreak cluster resulting from each introduction with larger circles representing more resulting downstream tips. Blue dashed line represents the linear best fit line using Pearson's correlation. Blue shaded region denotes the variability of the line and the resulting estimates from Pearson's correlation are shown in text above the shaded region.

Minor/trivial points

(For the future… line numbers are helpful for reviewers, assuming they are allowed.)
Some of these would help readers less familiar with the field to understand:

p. 6: State what discrete trait is being inferred.

We've added the following clarification:

To investigate the spread of mpox throughout the course of the epidemic across global regions, we employed a phylogeographic approach with an asymmetrical discrete trait model on 1004 publicly available MPXV sequences subsampled based on confirmed case counts (Fig. 2, Fig. S1) *in order to infer the global region of origin for all internal ancestral nodes.*

p.6-7: The predictors and the prediction approach (GLM) are stated, but not what is being predicted.

We've added the following clarification:

In order to analyze within-region transmission dynamics, improve robustness to sampling bias, and enhance inference via the joint integration of genomic and epidemiological metadata, we then employed an approximate structured coalescent (MASCOT) with a generalized linear model (GLM) approach with estimated prevalence and air passenger data as empirical predictors on 587 sequences *in order to infer the effective population size and migration rates within and between each region, respectively*

p. 7: State more clearly what is meant by 'equal temporal weighting'.

In order to remain within the required word limit, we included a reference in the main text on page 7 to the STAR Methods section under *MASCOT-GLM* which states:

For this analysis, we employed equal temporal subsampling to enrich for undersampled regions by randomly choosing a max of 11 sequences per region per calendar month via Augur filter (11), resulting in 587 included sequences. We chose an equal temporal subsampling scheme due to recent work showing that maximizing spatiotemporal diversity reduces bias in MASCOT-GLM (12)

p. 9: Fig. 4B doesn't give the meaning of the error bars.

We added the following clarification to the figure legend:

Error bars denote 95% HPD interval for the magnitude of predictor coefficient

p. 10: Define the units for effective population size.

We ensured that every figure with effective population size estimates denoted "*Neτ* (in years)" and additionally included the following in the legend of figure 3 with a reference to Bedford et al. (2011) for more information

the coalescent time scale depends on both effective population size Ne (number of effective individuals) and on generation time τ (years per generation), resulting in Neτ being a measure of coalescent time scale in years (13)

The reference to Fig. 3D ('driven by lack of genomic and case-based information at those time periods') doesn't look right, while the reference to Sup. Fig. 6 looks like it should be to Sup. Fig. 10.
p. 21: Typo: extra opening parenthesis in line after equation.

We have corrected the above errors on p10 and p21. Thank you for highlighting them.

p. 22: Perhaps I don't understand something, but I found this sentence baffling: 'A priori, we assumed that the effective population size at time t+1 is normally distributed with mean 0 and standard deviation σ, with σ being estimated'.

We rephrase this sentence to clarify, we now write:

"We assume the prior on the effective population size over time to be a Gaussian Markov random field (GMRF) and estimate the variance of the GMRF prior on the effective population size over time. We assume the GMRF prior for each state to have the same variance "

p. 24: I found the beginning of the page confusing. The cited reference (72) appears to be to the wrong paper (Musa et al rather than Ma), lowercase δ is undefined, and it was not clear to me where the equation came from or what model in Ma's paper it was relevant to.

Thank you for pointing this out. δ was erroneously included and actually refers to infectious period γ. This has been corrected. Additionally, we have updated the Ma reference to highlight that our equations were derived from "Example 4" in the Ma 2020 paper.