Inferring the multiplicity of founder variants initiating HIV-1 infection: a systematic review and individual patient data meta-analysis

James Baxter, Sarah Langhorne, Ting Shi, Damien C Tully, Ch Julián Villabona-Arenas, Stéphane Hué, Jan Albert, Andrew Leigh Brown, Katherine E Atkins

Summary

Background HIV-1 infections initiated by multiple founder variants are characterised by a higher viral load and a worse clinical prognosis than those initiated with single founder variants, yet little is known about the routes of exposure through which transmission of multiple founder variants is most probable. Here we used individual patient data to calculate the probability of multiple founders stratified by route of HIV exposure and study methodology.

Methods We conducted a systematic review and meta-analysis of studies that estimated founder variant multiplicity in HIV-1 infection, searching MEDLINE, Embase, and Global Health databases for papers published between Jan 1, 1990, and Sept 14, 2020. Eligible studies must have reported original estimates of founder variant multiplicity in people with acute or early HIV-1 infections, have clearly detailed the methods used, and reported the route of exposure. Studies were excluded if they reported data concerning people living with HIV-1 who had known or suspected superinfection, who were documented as having received pre-exposure prophylaxis, or if the transmitting partner was known to be receiving antiretroviral treatment. Individual patient data were collated from all studies, with authors contacted if these data were not publicly available. We applied logistic meta-regression to these data to estimate the probability that an HIV infection is initiated by multiple founder variants. We calculated a pooled estimate using a random effects model, subsequently stratifying this estimate across exposure routes in a univariable analysis. We then extended our model to adjust for different study methods in a multivariable analysis, recalculating estimates across the exposure routes. This study is registered with PROSPERO, CRD42020202672.

Findings We included 70 publications in our analysis, comprising 1657 individual patients. Our pooled estimate of the probability that an infection is initiated by multiple founder variants was 0·25 (95% CI 0·21–0·29), with moderate heterogeneity (Q=132·3, p<0·0001, I^2=64·2%). Our multivariable analysis uncovered differences in the probability of multiple variant infection by exposure route. Relative to a baseline of male-to-female transmission, the predicted probability for female-to-male multiple variant transmission was significantly lower at 0·13 (95% CI 0·08–0·20), and the probabilities were significantly higher for transmissions in people who inject drugs (0·37 [0·24–0·53]) and men who have sex with men (0·30 [0·33–0·40]). There was no significant difference in the probability of multiple variant transmission between male-to-female transmission (0·21 [0·14–0·31]), post-partum transmission (0·18 [0·03–0·57]), pre-partum transmission (0·17 [0·08–0·33]), and intra-partum transmission (0·27 [0·14–0·45]).

Interpretation We identified that transmissions in people who inject drugs and men who have sex with men are significantly more likely to result in an infection initiated by multiple founder variants, and female-to-male infections are significantly less probable. Quantifying how the routes of HIV infection affect the transmission of multiple variants allows us to better understand how the evolution and epidemiology of HIV-1 determine clinical outcomes.

Funding Medical Research Council Precision Medicine Doctoral Training Programme and a European Research Council Starting Grant.

Lancet Microbe 2023

Published Online January 12, 2023 https://doi.org/10.1016/S2666-5247(22)00327-5

Usher Institute (J Baxter BSc Hons, T Shi PhD, K E Atkins PhD) **and Institute of Evolutionary Biology** (Prof A Leigh Brown PhD), **The University of Edinburgh, Edinburgh, UK; Department of Infectious Disease Epidemiology, Faculty of Epidemiology and Population Health** (S Langhorne MSc, C J Villabona-Arenas ScD, D C Tully PhD, S Hué PhD, K E Atkins) **and Centre for Mathematical Modelling of Infectious Diseases** (C J Villabona-Arenas, D C Tully, S Hué, K E Atkins)**, London School of Hygiene & Tropical Medicine, London, UK; Department of Microbiology, Tumor and Cell Biology, Karolinska Institutet, Stockholm, Sweden** (Prof J Albert MD)**; Department of Clinical Microbiology, Karolinska University Hospital, Stockholm, Sweden** (Prof J Albert)

Correspondence to: Mr James Baxter, Usher Institute, The University of Edinburgh, Edinburgh EH9 3FL, UK **james.baxter@ed.ac.uk**

Introduction

Transmission of HIV-1 results in a dramatic reduction in genetic diversity, with a large proportion of infections initiated by a single founder variant.[1,2] An appreciable minority of infections, however, appear to be the result of multiple founder variants simultaneously initiating infection after a single exposure.[3] Importantly, these infections caused by multiple founder variants are associated with elevated set point viral load and faster CD4-positive T lymphocyte decline.[4–7]

HIV-1 infections initiated via different routes of exposure are subject to different virological, cellular, and physiological environments, which probably influence the probability of acquiring infection.[8–10] For example, the per-act probability of transmission upon exposure is six times higher for transmission between people who

Research in context

Evidence before this study

Most HIV-1 infections are initiated by a single, genetically homogeneous founder variant. Infections initiated by multiple founders, however, are associated with a significantly faster decline of CD4-positive T cells in untreated individuals, ultimately leading to an earlier onset of AIDS. Through our systematic search of MEDLINE, Embase, and Global Health databases for papers published between Jan 1, 1990, and Sept 14, 2020, we identified 82 studies that classify the founder variant multiplicity of early HIV infections. Estimates of the probability that infection was initiated by multiple founders vary widely across these studies, and present inconsistent conclusions as to the risk of multiple founder infection across routes of exposure. To better understand the relationship between the probability of infection and the probability of multiple founders, it is important to clarify the extent to which route of exposure determines the probability that infection is initiated by multiple founders.

Added value of this study

Comparing the probability of multiple founder infection across routes of exposure has been challenging due to generational changes in methodology across studies and underlying heterogeneity in the populations studied. By adjusting for sources of study-level and patient-level heterogeneity in our multivariable meta-regression fitted to individual patient data from 70 studies, we were able to estimate the probability that infection is initiated by multiple founders by route of HIV exposure. We uncovered differences in the probability that an infection is initiated by multiple founder variants by exposure route, with people who inject drugs and men who have sex with men at greater risk of infections founded by multiple variants. Further, by systematically collating all available data, we were able to adjust our estimates of founder variant multiplicity for methodological differences between studies that have made previous comparisons difficult.

Implications of all the available evidence

Because HIV-1 infections initiated by multiple founders are associated with a poorer prognosis, determining whether the route of exposure affects the probability with which infections are initiated by multiple variants facilitates an improved understanding of how the evolution and epidemiology of HIV-1 determine clinical progression. Our results identify that transmissions in people who inject drugs and men who have sex with men are significantly more likely to result in an infection initiated by multiple founder variants compared with male-to-female transmissions. This finding reiterates the need for focused public health programmes that reduce the burden of HIV-1 in these at-risk groups.

inject drugs and up to 18 times higher for transmission between men who have sex with men than for heterosexual transmission.[11]

Despite these differences in the probability of HIV-1 acquisition by route of exposure, there is currently no consensus about whether the route of exposure determines the probability that infection is initiated by multiple founder variants. Differences in selection pressure during transmission have been observed between sexual exposure routes, with less selection occurring during sexual transmission from males to females than vice versa, and less selection during transmission in men who have sex with men relative to heterosexual exposure overall.[12,13] Less selection should lead to more opportunities for infections initiated with more founder variants. Studies quantifying the number of founder variants are, however, inconsistent with these findings, which might be due to differences in methodology and study population.[3,12,14,15] Moreover, although acquisition risk during sexual transmission is known to be elevated during conditions that increase mucosal inflammation and compromise mucosal integrity, there is no consistent evidence that transmissions in people who inject drugs, which bypass mucosal barriers altogether, are associated with a higher probability of founder variant initiation.[16,17] To estimate the role of exposure route on the acquisition of multiple HIV-1 founder variants, we conducted a meta-regression of all available individual patient data, accounting for heterogeneity across methodology and study population.

Methods

Search strategy and eligibility criteria

In this systematic review and meta-analysis we searched MEDLINE, Embase, and Global Health databases for papers published between Jan 1, 1990, and Sept 14, 2020; a full list of search terms is listed in the appendix (pp 2–6). To be included, studies must have reported original estimates of founder variant multiplicity in people with acute or early HIV-1 infections, be written in English, and document ethical approval. No restrictions were placed on study design, geographical location, or age of participants. Studies were excluded if they did not distinguish between single and multiple founder variants, if they did not detail the methods used, or if the study was conditional on having identified multiple founders. Additionally, studies were excluded if they solely reported data concerning people living with HIV-1 who had known or suspected superinfection, who were documented as having received pre-exposure prophylaxis, or if the transmitting partner was known to be receiving antiretroviral treatment. Studies were screened independently by SL and JB. Reviewers were blinded to study authorship during the title and abstract screens, and full-text reviews were conducted independently

See Online for appendix

before a consensus was reached, consulting other co-authors when necessary. This review conforms to PRISMA guidelines (appendix pp 7–10).

Data extraction

Individual patient data (IPD) were collated from all studies, with authors contacted if these data were not readily available. Studies were excluded from further analysis if IPD could not be obtained. Only individuals for whom a route of exposure was known were included. Additionally, we removed any entries for individuals with known or suspected superinfection, who were receiving pre-exposure prophylaxis or for whom the transmitting partner was known to be receiving antiretroviral therapy. For the base case dataset, we recorded whether an infection was initiated by one or multiple variants and eight predetermined covariates to be considered in the multivariable meta-regression.

The first covariate was the route of exposure: female to male, male to female, men who have sex with men, pre-partum transmission, intra-partum transmission, and post-partum transmission, or people who inject drugs. Two further categories were included for heterosexual and mother-to-child exposures whereby the direction or timing respectively were not disclosed (heterosexual undisclosed or mother-to-child undisclosed).

The second covariate was the quantification method whereby methodological groupings were defined by the properties of each approach, resulting in six levels: phylogenetic (using source and recipient sequences), phylogenetic (using recipient sequences only), haplotype, distance, model, or molecular (appendix p 4).

The third covariate was the HIV subtype: infecting subtypes were classed as either a canonical geographically delimited subtype (A–D, F–H, J, and K), a circulating recombinant form (CRF), or recombinant (when a putative recombinant was identified but not designated a CRF).[18,19]

The fourth covariate was delay between infection and sampling. For sexual exposure or people-who-inject-drugs exposure, the delay was classified as either less than or equal to 21 days if the patient was seronegative at time of sampling (Fiebig stages I–II) or more than 21 days if the patient was seropositive (Fiebig stages III-VI). For mother-to-child infections, if infection was confirmed at birth, or within 21 days of birth, the delay was classified as either less than or equal to 21 days. A positive mRNA or antibody test reported after this period was classified as a delay of greater than 21 days.

The fifth covariate was the number of genomes analysed per participant. For studies that use single genome amplification (SGA), this was the number of consensus genomes obtained.

The sixth covariate was the genomic region analysed. The region was classified as envelope (*env*), *pol*, *gag*, or near full length genome (NFLG).

The seventh covariate was the alignment length analysed. The length was measured in base pairs, discretised to the nearest 250 bp, 500 bp, 1000 bp, 2000 bp, 4000 bp, 8000 bp, and NFLG intervals (around 9000 bp).

Finally, the eighth covariate was the use of SGA to generate viral sequences. A binary classification was used to characterise whether the viral genomic data were generated using SGA. SGA mitigates the risk of Taq-polymerase mediated template switching, nucleotide misincorporation, or unequal amplicons resampling encountered in regular bulk or near endpoint PCR amplification.[20–22]

If information from covariates three to eight was missing or could not be inferred from the study, we classified its value as unknown. We excluded covariate levels for which there were fewer than six data points. For our main analysis, we removed repeat measurements for the same individual, and used only those from the earliest study or, where the results of different methods were reported by the same study, the conclusive method used for each individual. Further details on covariate selection are in the appendix (pp 2–4).

Statistical analysis

We calculated a pooled estimate of the probability of multiple founder variant infection using a one-step generalised linear mixed model, assuming an exact binomial distribution with a normally distributed random effect on the intercept for within-study clustering and fitted by approximate maximum likelihood.[23] Heterogeneity was measured in terms of τ^2, the between-study variance; I^2, the percentage of variance attributable to study heterogeneity; and Cochran's Q, an indicator of larger variation between studies than of participants within studies.[24] Publication bias was assessed using funnel plots and Egger's regression test.[25] All analyses were conducted in R 4.1.2 using packages metafor 3.0-2 for pooling and lme4 1.1-27.1 for the meta-regression.[26–28]

Pooled estimates obtained through a one-step approach are usually congruent with the canonical two-step meta-analysis model; however discrepancies might arise due to differences in weighting schemes, specification of the intercept, or estimation of residual variances.[29] We compared the results from our one-step model with a two-step binomial-normal model to confirm our estimates were consistent. We also performed seven sensitivity analyses to test the robustness of our pooled estimate: (1) iteratively excluding single studies, (2) excluding studies that contained fewer than ten participants, (3) setting variable thresholds of the number of genomes per participant, (4) excluding studies that consisted solely of single founder infections, (5) excluding IPD that did not use SGA, (6) including only those data that matched a gold-standard methodology of haplotype-based methods and envelope genomic region analysis, and (7) an assessment of the effect of vaccine breakthrough and sequencing technologies. To validate our down-sampling method that used only the most recent study for repeated

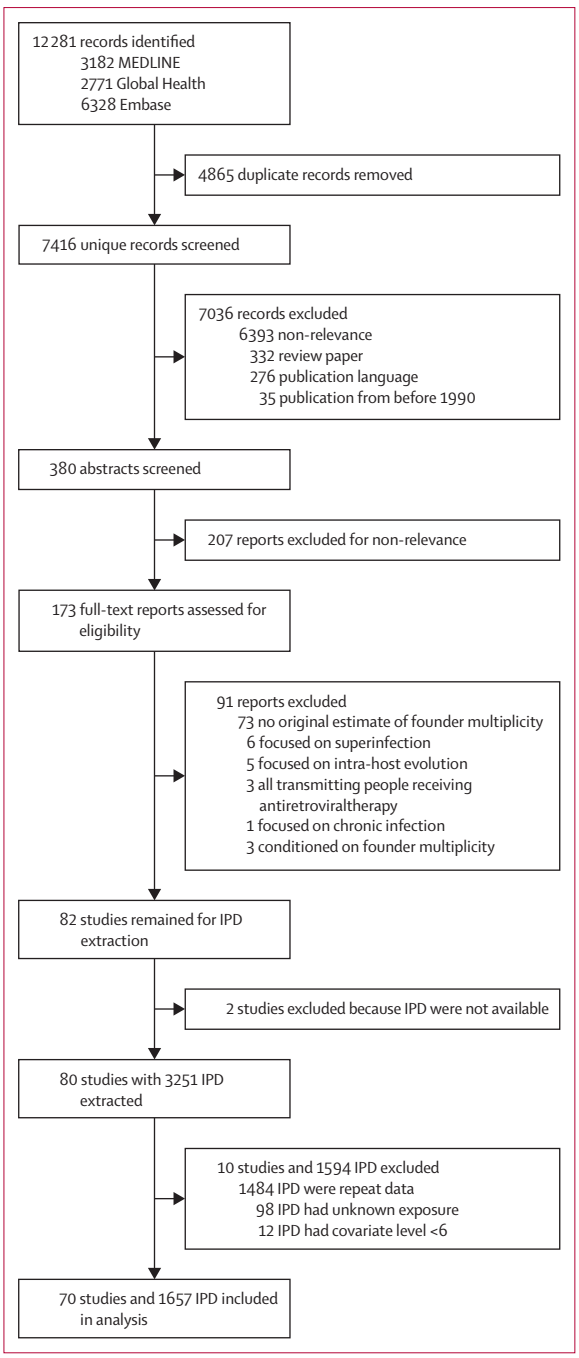

Figure 1: **PRISMA flowchart for the systematic literature search and the application of exclusion criteria for the IPD meta-analysis**
IPD=individual patient data.

individual data, we calculated a distribution of pooled estimates by refitting the pooling models to 1000 datasets, each containing one data point per individual sampled at random from an individual's possible measurements.

We extended our one-step model by conducting a univariable meta-regression with each covariate contributing a fixed effect and assuming normally distributed random effects of publication. We extended this model to a multivariable analysis. Fixed effects were selected according to a keep-it-maximal principle, in which covariates were only removed to facilitate a non-singular fit and to prevent multicolinearity.[30] For both univariable and multivariable analyses, we defined our reference case as male-to-female transmission and evaluated through a gold-standard methodology of haplotype-based methods, analysis of the envelope genomic region, and a sampling delay of less than or equal to 21 days. We report odds ratios (ORs) and stratified model estimates of the proportion of infections initiated by multiple founders across each route of exposure with all other covariates held at their reference case values. Bootstrapped 95% CIs were calculated for both ORs and stratified model estimates. We performed four sensitivity analyses to test the robustness of the selected multivariable meta-regression model: (1) iteratively excluding single studies, (2) excluding studies that contained fewer than ten participants, (3) excluding studies that consisted solely of single founder infections, and (4) excluding IPD that did not use SGA. The resampling sensitivity analysis was repeated on our selected multivariable model as described earlier. Further details are in the appendix (pp 4–6). The protocol for this study was registered with PROSPERO (CRD42020202672).

Role of the funding source

The funders of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the report.

Results

Our search found 7416 unique papers, of which 7334 were excluded, leaving 82 studies remaining (figure 1).[3,5–7,12,14–17,22,31–104] Of these studies, we extracted IPD from 80 studies, comprising 3251 data points. The 80 selected studies from which IPD were collated, were published between 1992 and 2020. Of the 3251 data points extracted, 1484 were excluded from our base case dataset to avoid repeated measurements arising either between different studies that analysed the same individuals (resulting in the exclusion of five studies), or from repeat analysis of individuals within the same study. After excluding 98 participants for whom the route of exposure was unknown and 12 participants for whom one or more of the covariate values pertained to a covariate level that did not meet the minimum number (six) of observations across all participants, the base case dataset for our analysis comprised estimates from 1657 unique patients across 70 studies.

Our base case dataset includes a median of 13 participants per study (IQR 7–35) and represents infections associated with heterosexual transmission (696 [42·0%] of 1657), transmission in men who have sex with men (621 [37·5%]), mother-to-child transmission (234 [14·1%]), and transmission in people who inject

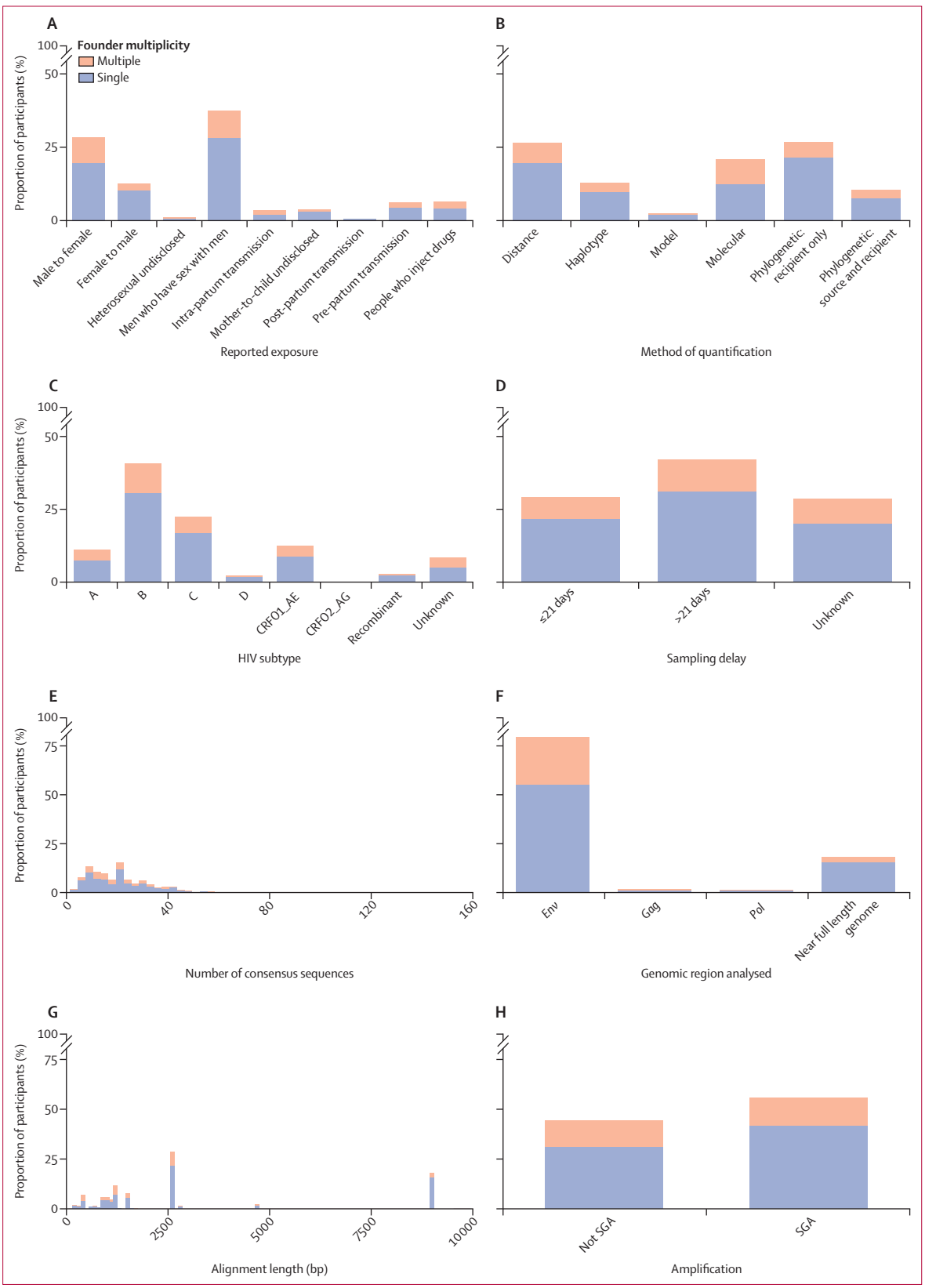

Figure 2: **Individual patient data characteristics from the included studies that were tested for inclusion as fixed effects in the multivariable meta-regression model**
SGA=single genome amplification.

	Male-to-female transmission	Female-to-male transmission	Heterosexual unknown transmission	Men-who-have sex-with-men transmission	People-who-inject-drugs transmission	Pre-partum transmission	Intra-partum transmission	Post-partum transmission	Unknown mother-to-child transmission
Number of studies	32 (39)	25 (30)	3 (4)	28 (34)	12 (13)	7 (7)	7 (7)	1 (1)	6 (6)
Number of participants	471 (601)	208 (319)	17 (22)	621 (812)	106 (116)	104 (104)	57 (57)	11 (11)	62 (62)
Number of participants with multiple founder variant estimated	147 (188)	39 (61)	5 (7)	154 (205)	38 (45)	31 (31)	25 (25)	2 (2)	12 (12)
Quantification methods									
Distance	65 (14%)	67 (32%)	2 (12%)	80 (13%)	1 (1%)	..	..	..	..
Haplotype	105 (22%)	73 (35%)	4 (24%)	139 (22%)	63 (60%)	..	..	..	54 (87%)
Model	9 (2%)	2 (1%)	..	10 (2%)	14 (13%)	2 (2%)	2 (4%)	..	..
Molecular	161 (34%)	26 (13%)	..	27 (4%)	9 (8%)	92 (88%)	32 (56%)	..	..
Phylogenetic: recipient only	99 (21%)	25 (12%)	..	305 (49%)	14 (13%)	..	..	..	..
Phylogenetic: source and recipient	32 (7%)	15 (7%)	11 (65%)	60 (10%)	5 (5%)	10 (10%)	23 (40%)	11 (100%)	8 (13%)
Genomic regions analysed									
Env	437 (93%)	179 (86%)	15 (88%)	351 (57%)	101 (95%)	103 (99%)	57 (100%)	11 (100%)	61 (98%)
Gag	14 (3%)	8 (4%)	..	..	..	1 (1%)	..	..	1 (2%)
Pol	1 (<0·5%)	3 (1%)	..	13 (2%)	1 (1%)	..	..	..	..
Near full length genome	19 (4%)	18 (9%)	2 (12%)	257 (41%)	4 (4%)	..	..	..	..
Median sequences per participant	21 (12–36)	21 (12–38)	19 (11–56)	16 (4–40)	23 (20–23)	32 (19–39)	20 (6–25)	..	38 (34–42)

Data are n (N) where N are quantities before removal of repeat participants, n (%), or median (IQR). The denominator for percentages is the number of participants. Percentages are rounded to nearest whole number.

Table 1: **Summary of individual and study characteristics in our base case dataset by transmission route**

drugs (106 [6·4%]; figure 2; table 1; appendix pp 11–17). Our dataset spans all major subtypes and captures the diversity of the HIV epidemic (figure 2; appendix p 18). Across the base case dataset, 618 (37·3%) estimates were inferred using phylogenetic methods, 438 (26·4%) using haplotype-based methods, 347 (20·9%) using molecular methods, 215 (13·0%) using distance-based methods, and 39 (2·4%) using model-based methods (figure 2). Data from 1315 (79·4%) participants were analysed from the *env* genomic region, 300 (18·1%) from NFLG, 24 (1·4%) from the *gag* region, and 18 (1·1%) from the *pol* region.

Our binomial generalised linear mixed model pooled estimated the probability that an infection is initiated by multiple founder variants as 0·25 (95% CI 0·21–0·29), identifying moderate heterogeneity (Q=132·3, p<0·0001, I^2=64·2%). Visual inspection of a funnel plot and a non-significant Egger's test (t=–0·7495, df=55, p=0·46) were consistent with an absence of publication bias (appendix p 25). Sensitivity analyses revealed the pooled estimate was robust to the choice of model, the inclusion of estimates from repeat participants, and to the exclusion of studies that contained fewer than ten participants or did not report any participants with multiple founder infections (appendix p 19). Although restricting the analysis to participants for whom a large (>28) number of sequences were analysed did not change the pooled estimate (0·26 [95% CI 0·20–0·34]), restricting the analysis to those individuals with fewer than 11 sequences reduced the estimate to 0·21 (0·17–0·25; appendix p 20). Analysing only data that matched our gold-standard study methodology slightly increases the pooled estimate (0·28 [0·22–0·35]; appendix p 20). We did not identify any studies or risk groups that individually influenced the pooled estimate significantly (appendix pp 21–22). A pooled estimate subgroup analysis of placebo and vaccine participants from studies for which vaccination status was available revealed no discernible influence of trial group (appendix p 23). Likewise, no discernible difference was identified between sequencing technologies on the pooled estimate (appendix p 24).

We first extended our binomial generalised linear mixed model with univariable fixed effects. Relative to a reference exposure route of male-to-female transmission, we found significantly lower odds of female-to-male transmission being initiated by multiple founder variants (OR 0·53 [95% CI 0·33–0·85]), but other exposure routes were not significantly different (table 2). The univariable analyses also indicate significantly lower odds of identifying multiple founder variants when the NFLG was analysed (0·38 [0·19–0·68]) relative to the envelope genomic region, and molecular methods resulted in significantly greater odds (1·93 [1·02–3·45]) relative to haplotype methods.

Next, we used a multivariable model to calculate the probability of multiple founder variants across the seven routes of exposure controlling for method, genomic

region, and sampling delay (appendix p 27; table 2). A satisfactory fit was confirmed by inspection of binned residuals superimposed over 95% CIs (appendix p 28). Model estimated probabilities were calculated with respect to our gold-standard methodology (figure 3; appendix p 26). Compared with a male-to-female transmission probability of 0·21 (95% CI 0·14–0·31), we found that female-to-male transmissions were less likely to be initiated by multiple founders, with a probability of 0·13 (0·08–0·21; OR 0·55 [95% CI 0·34–0·88]; figure 3; appendix p 26). Conversely, transmissions were more likely to be initiated by multiple founders for people who inject drugs (probability 0·37 [0·24–0·53]; OR 2·18 [1·11–3·89]) and men who have sex with men (probability 0·30 [0·22–0·40]; OR 1·61 [1·00–2·34]), compared with male-to-female transmission. Stratifying mother-to-child transmissions by the putative timing of infection, we calculated pre-partum transmissions were initiated by multiple founders with a probability of 0·17 (0·08–0·33), post-partum transmissions with a probability of 0·18 (0·03–0·57), and intra-partum transmissions with a probability of 0·27 (0·14–0·45). No mother-to-child transmission group was significantly different from the estimated probability that infection is initiated by multiple founders in male-to-female transmissions (table 2).

We inferred the accuracy of different methods by comparing their estimated probability of multiple founder variants to a gold-standard methodological reference scenario of haplotype-based methods on the envelope genomic region with individuals with less than or equal to 21 days between infection and sampling (table 2). Our analysis indicates using model-based methods underestimates the chance of multiple founder variants (OR 0·36 [95% CI 0·09–0·87]), and using molecular methods results in an overestimation (2·05 [1·09–3·53]). Compared with the envelope genomic region, analysis of NFLG fragments probably underestimates the proportion of multiple founder infections (0·31 [0·13–0·62]). No significant difference was revealed between the odds of multiple founder infections sampled less than or on 21 days and at least 21 days after the estimated date of infection (1·16 [0·78–1·65]). Our sensitivity analyses revealed the ORs calculated using the univariable and multivariable models are robust to inclusion of data from repeated participants, to the exclusion of studies that contained fewer than ten participants, of studies that consisted solely of single founder infections, and of individual data that did not use SGA (appendix p 29).

	Univariable meta-regression model		Multivariable meta-regression model	
	OR (95% CI)	p value	OR (95% CI)	p value
Reported exposure				
Male to female	1 (ref)	..	1 (ref)	..
Female to male	0·53 (0·33–0·85)	0·0063	0·55 (0·34–0·90)	0·011
Heterosexual unknown	1·81 (0·40–5·91)	0·34	1·72 (0·25–5·24)	0·36
Men who have sex with men	1·33 (0·83–2·03)	0·24	1·61 (1·00–2·34)	0·023
Pre-partum transmission	1·26 (0·54–2·65)	0·59	0·76 (0·38–1·59)	0·48
Intra-partum transmission	1·87 (0·81–4·02)	0·15	1·34 (0·58–2·86)	0·46
Post-partum transmission	0·73 (0·01–3·56)	0·77	0·79 (0·00–3·58)	0·79
Unknown mother-to-child transmission	1·23 (0·45–3·50)	0·70	0·79 (0·29–2·23)	0·64
People who inject drugs	2·08 (0·91–4·15)	0·051	2·18 (1·17–3·89)	0·018
Quantification method				
Haplotype	1 (ref)	..	1 (ref)	..
Distance	0·76 (0·35–1·58)	0·44	1·46 (0·76–2·86)	0·25
Model	0·53 (0·09–1·39)	0·27	0·36 (0·09–0·87)	0·057
Molecular	1·93 (1·02–3·45)	0·026	2·05 (1·09–3·53)	0·018
Phylogenetic: recipient only	0·72 (0·42–1·24)	0·23	0·83 (0·48–1·54)	0·47
Phylogenetic: source and recipient	0·90 (0·49–1·57)	0·73	0·95 (0·52–1·92)	0·85
Genomic region				
env	1 (ref)	..	1 (ref)	..
Near full length genome	0·38 (0·19–0·63)	0·0026	0·31(0·13–0·62)	0·0003
gag	1·13 (0·22–4·51)	0·86	2·14(0·45–6·80)	0·22
pol	0·31 (0·00–1·36)	0·17	0·31(0·00–1·13)	0·16
Sampling delay				
≤21 days	1 (ref)	..	1 (ref)	..
>21 days	1·09 (0·74–1·61)	0·63	1·16 (0·78–1·65)	0·43
Unknown	1·42 (0·84–2·72)	0·20	1·39 (0·81–2·43)	0·22

OR=odds ratio.

Table 2: **ORs that an HIV–1 infection is initiated by multiple founder variants, inferred from fixed effects coefficients from the univariable and multivariable meta-regression model**

Discussion

Using data from 70 published studies, we estimated that a quarter of HIV-1 infections are initiated by multiple founder variants. When controlling for different methodologies across studies, the probability that an infection is initiated by multiple founders decreased from 0·21 for male-to-female infections to 0·13 for female-to-male infections, but increased to 0·30 for infections in men who have sex with men and to 0·37 for infections in people who inject drugs. Further, we found that model-based methods, representing a group of approaches that determine founder multiplicity by comparing the observed distribution of diversity with that expected under neutral exponential outgrowth from single variant transmission, were less likely to identify multiple founder infections whereas molecular methods overestimated. Together these results suggest that while the exposure route probably influences the number of founder variants, previous comparison has been difficult due to different study methodologies.

Our pooled estimate is consistent with the seminal study by Keele and colleagues,[3] who found that 24 (23·5%) of 102 participants had infections initiated by multiple founders. Our stratified predicted probabilities, however, were marginally higher than those of previous studies. A nine-study meta-analysis of 354 participants found the probability of HIV infections being initiated by multiple

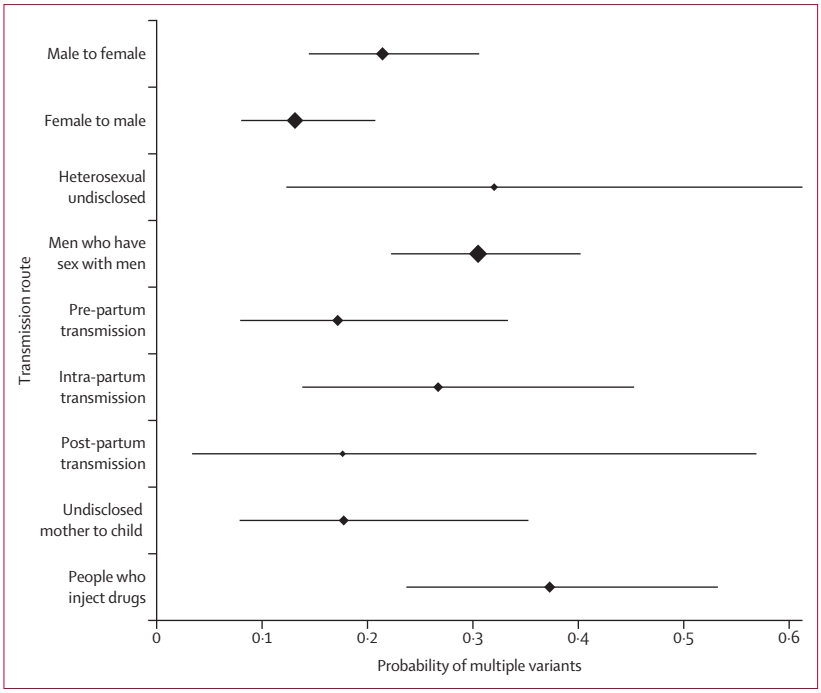

Figure 3: **Model estimated probabilities of an infection being initiated by multiple founder variants, stratified by the route of exposure**

founders to be 0·34 in people who inject drugs versus 0·37 (95% CI 0·24–0·53) in our study and 0·25 in men who have sex with men versus 0·30 (0·22–0·40) in our study.[12] An earlier meta-analysis of five studies and 235 participants also found HIV infections in people who inject drugs were at significantly greater odds than heterosexual infections of being initiated by multiple founders, with the frequency of founder variant multiplicity increasing 3-fold, while a smaller, non-significant 1·5-fold increase was observed with respect to transmissions in men who have sex with men.[16] In both instances, these studies restricted participants so that the methodology in estimating founder variant multiplicity was consistent. Here, we were able to leverage individual-level data to control for methodological sources of heterogeneity across studies.

Between the conclusion of our analysis and publication, one additional study presented novel estimations of founder variant multiplicity.[105] Using diversity, haplotype, and phylogenetic methods, Balinda and colleagues[105] found 27% (seven of 26) of infections were initiated by multiple founders, consistent with our pooled probability estimate of 0·25 (95% CI 0·21–0·29).

Across sexual transmission routes, the probability of multiple founder variants is positively, albeit weakly, associated with an increase in the risk of transmission given exposure.[11] Nonetheless, the probability that infection is initiated by multiple founders remains remarkably consistent. For example, although male-to-male exposures might be up to 18 times more likely to result in transmission than male-to-female exposures, we calculated a 1·6-fold increase in the risk of multiple founders.[11] Previously, Thompson and colleagues reconciled the low probability of acquisition with the relatively high probability of multiple founders by assuming only a fraction of exposures occur in environments conducive for transmission.[106] In sexual transmission, for example, a conducive environment could arise through epithelial damage after ulceration or microtrauma that enhances translocation of viral particles or drives inflammation that propagates recruitment of permissive target cells.[8] The relatively consistent probability of multiple variant transmission across different routes of exposure suggests the presence of selection at transmission, which has been observed previously.[13]

Our analysis has some limitations. First, our classification of founder variant multiplicity is determined by the individual studies, but explicitly defining a founder variant remains challenging. Recent studies have suggested a continuum of genotypic diversity exists, rather than discrete variants, that gives rise to distinct phylogenetic diversification trajectories and might not be reflected by a binary classification.[34,70] Although a threshold is specified for distance-based methods, this often varies between publications.[107,108] For example, both Keele and colleagues[3] and Li and colleagues[15] analysed the diversity of the envelope protein, but Keele and colleagues classified populations with less than 0·47% diversity as homogeneous and Li and colleagues included samples up to 0·75%. The distinction between single and multiple founder variants might further be blurred by recombination and hypermutation.[109,110] Our finding that the analysis of NFLGs was associated with a significant decrease in the odds of multiple founders suggests earlier studies that rely on smaller, highly variable, fragments of envelope might have overestimated the frequency of infections initiated by multiple founder variants. Similarly, our sensitivity analyses revealed a subtle correlation between the number of genomes analysed and the probability of observing multiple founder variants, pointing to the possibility that using too few genomes could limit the chance of observing multiple founders.

Second, we acknowledge that some heterogeneity associated with our estimates is encapsulated within the classification of route of exposure. Relying on self-reported route of exposure might bias our results if misclassification occurs systematically across studies. Similarly, insufficient data were available to properly consider risk factors such as genital ulceration, early stage of disease in the transmitting person, or receptive anal intercourse. These risk factors might confound or mediate any association between the exposure type and the probability of multiple founder variants, potentially hindering a deeper mechanistic understanding as to the risk factors underpinning founder variant multiplicity.[11] Also, under the hypothesis that the proportion of infections initiated by multiple founders varies by transmission route, our point estimate will be influenced

by their relative proportions in our dataset. Globally, it is estimated that 70% of infections are transmitted heterosexually, compared with 42·2% in our dataset.[111] Our point estimate should be considered a summary of the published data over the course of the HIV-1 epidemic, and not a global estimate at any fixed point in time.

Finally, for several covariates, the bootstrapped CIs are wide and might lead to some uncertainty. These CIs are a product of small sample sizes for some observations, combined with the random effect of publication used in the meta-regression.

This systematic review and meta-analysis shows that infections initiated by multiple founders account for a quarter of HIV-1 infections across major routes of transmission. We found that transmissions involving people who inject drugs and men who have sex with men are significantly more likely to be initiated by multiple founder variants, than male-to-female infections, whereas female-to-male infections are significantly less probable. Given the positive association between multiple variant infections and a faster decline in CD4-positive T cells in untreated individuals, there is a clinical imperative to improve our understanding of the risk factors influencing both transmission and multiple founder infections. This study, for the first time, consolidates the existing body of research on the probability that HIV infection is initiated by multiple variants. Providing these estimates might precipitate further investigation to understand the evolutionary epidemiology that governs the clinical picture of HIV transmission.

Contributors

KEA conceived the study. JB, SL, DCT, and KEA designed the study. JB and SL extracted the data. JB analysed the data. JB and KEA verified the data. All authors interpreted the data. JB and KEA drafted the manuscript, with critical revisions from all authors. KEA, SH, and ALB supervised the study. All authors had full access to all the data in the study and had final responsibility for the decision to submit for publication.

Declaration of interests

We declare no competing interests.

Data sharing

Code and individual patient data used in this study are publicly available on GitHub.

Acknowledgments

JB was supported by the Medical Research Council Precision Medicine Doctoral Training Grant (2016–21; MR/N013166/1); CJV-A and KEA were funded by a European Research Council Starting Grant (award number 757688) awarded to KEA. We are grateful to Morgane Rolland for agreeing to share additional individual patient data with the authors to complete this study. We thank the four anonymous reviewers for their helpful feedback.

References

1 Zhu T, Mo H, Wang N, et al. Genotypic and phenotypic characterization of HIV-1 patients with primary infection. *Science* 1993; **261:** 1179.

2 Zhang LQ, MacKenzie P, Cleland A, Holmes EC, Brown AJ, Simmonds P. Selection for specific sequences in the external envelope protein of human immunodeficiency virus type 1 upon primary infection. *Journal of Virology* 1993; **67:** 3345.

3 Keele BF, Giorgi EE, Salazar-Gonzalez JF, et al. Identification and characterization of transmitted and early founder virus envelopes in primary HIV-1 infection. *Proc Natl Acad Sci USA* 2008; **105:** 7552–57.

4 Sagar M, Lavreys L, Baeten JM, et al. Infection with multiple human immunodeficiency virus type 1 variants is associated with faster disease progression. *J Virol* 2003; **77:** 12921–26.

5 Cornelissen M, Pasternak AO, Grijsen ML, et al. HIV-1 dual infection is associated with faster CD4+ T-cell decline in a cohort of men with primary HIV infection. *Clin Infect Dis* 2012; **54:** 539–47.

6 Janes H, Herbeck JT, Tovanabutra S, et al. HIV-1 infections with multiple founders are associated with higher viral loads than infections with single founders. *Nat Med* 2015; **21:** 1139–41.

7 Macharia GN, Yue L, Staller E, et al. Infection with multiple HIV-1 founder variants is associated with lower viral replicative capacity, faster CD4+ T cell decline and increased immune activation during acute infection. *PLoS Pathog* 2020; **16:** e1008853.

8 Kariuki SM, Selhorst P, Ariën KK, Dorfman JR. The HIV-1 transmission bottleneck. *Retrovirology* 2017; **14:** 22.

9 Joseph SB, Swanstrom R, Kashuba ADM, Cohen MS. Bottlenecks in HIV-1 transmission: insights from the study of founder viruses. *Nat Rev Microbiol* 2015; **13:** 414–25.

10 Talbert-Slagle K, Atkins KE, Yan K-K, et al. Cellular superspreaders: an epidemiological perspective on HIV infection inside the body. *PLoS Pathog* 2014; **10:** e1004092.

11 Patel P, Borkowf CB, Brooks JT, Lasry A, Lansky A, Mermin J. Estimating per-act HIV transmission risk: a systematic review. *AIDS* 2014; **28:** 1509–19.

12 Tully DC, Ogilvie CB, Batorsky RE, et al. Differences in the selection bottleneck between modes of sexual transmission influence the genetic composition of the HIV-1 founder virus. *PLoS Pathog* 2016; **12:** e1005619.

13 Carlson JM, Schaefer M, Monaco DC, et al. HIV transmission. Selection bias at the heterosexual HIV-1 transmission bottleneck. *Science* 2014; **345:** 1254031.

14 Haaland RE, Hawkins PA, Salazar-Gonzalez J, et al. Inflammatory genital infections mitigate a severe genetic bottleneck in heterosexual transmission of subtype A and C HIV-1. *PLoS Pathog* 2009; **5:** e1000274.

15 Li H, Bar KJ, Wang S, et al. High multiplicity infection by HIV-1 in men who have sex with men. *PLoS Pathog* 2010; **6:** e1000890.

16 Bar KJ, Li H, Chamberland A, et al. Wide variation in the multiplicity of HIV-1 infection among injection drug users. *J Virol* 2010; **84:** 6241–47.

17 Masharsky AE, Dukhovlinova EN, Verevochkin SV, et al. A substantial transmission bottleneck among newly and recently HIV-1-infected injection drug users in St Petersburg, Russia. *J Infect Dis* 2010; **201:** 1697–702.

18 Robertson DL, Anderson JP, Bradac JA, et al. HIV-1 nomenclature proposal. *Science* 2000; **288:** 55–56.

19 Archer J, Robertson DL. Understanding the diversification of HIV-1 groups M and O. *AIDS* 2007; **21:** 1693–700.

20 Meyerhans A, Vartanian J-P, Wain-Hobson S. DNA recombination during PCR. *Nucleic Acids Res* 1990; **18:** 1687–91.

21 Simmonds P, Balfe P, Peutherer JF, Ludlam CA, Bishop JO, Brown AJ. Human immunodeficiency virus-infected individuals contain provirus in small numbers of peripheral mononuclear cells and at low copy numbers. *J Virol* 1990; **64:** 864–72.

22 Salazar-Gonzalez JF, Bailes E, Pham KT, et al. Deciphering human immunodeficiency virus type 1 transmission and early envelope diversification by single-genome amplification and sequencing. *J Virol* 2008; **82:** 3952–70.

23 Riley RD, Legha A, Jackson D, et al. One-stage individual participant data meta-analysis models for continuous and binary outcomes: comparison of treatment coding options and estimation methods. *Stat Med* 2020; **39:** 2536–55.

24 Borenstein M, Hedges LV, Higgins JPT, Rothstein HR. Introduction to meta-analysis. Chichester, UK: John Wiley & Sons, 2011.

25 Egger M, Davey Smith G, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. *BMJ* 1997; **315:** 629–34.

26 R Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing, 2021.

27 Bates D, Mächler M, Bolker B, Walker S. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 2015; **67:** 1–48.

For **code and individual patient data** see https://github.com/J-Baxter/foundervariantsHIV_sysreview

28 Viechtbauer W. Conducting meta-analyses in R with the metafor package. *J Stat Softw* 2010; **36:** 1–48.

29 Burke DL, Ensor J, Riley RD. Meta-analysis using individual participant data: one-stage and two-stage approaches, and why they may differ. *Stat Med* 2017; **36:** 855–75.

30 Barr DJ, Levy R, Scheepers C, Tily HJ. Random effects structure for confirmatory hypothesis testing: keep it maximal. *J Mem Lang* 2013; **68:** 255–78.

31 Sagar M, Kirkegaard E, Long EM, et al. Human immunodeficiency virus type 1 (HIV-1) diversity at time of infection is not restricted to certain risk groups or specific HIV-1 subtypes. *J Virol* 2004; **78:** 7279–83.

32 Abrahams M-R, Anderson JA, Giorgi EE, et al. Quantitating the multiplicity of infection with human immunodeficiency virus type 1 subtype C reveals a non-Poisson distribution of transmitted variants. *J Virol* 2009; **83:** 3556–67.

33 Sagar M, Kirkegaard E, Lavreys L, Overbaugh J. Diversity in HIV-1 envelope V1-V3 sequences early in infection reflects sequence diversity throughout the HIV-1 genome but does not predict the extent of sequence diversity during chronic infection. *AIDS Res Hum Retroviruses* 2006; **22:** 430–37.

34 Lewitus E, Rolland M. A non-parametric analytic framework for within-host viral phylogenies and a test for HIV-1 founder multiplicity. *Virus Evol* 2019; **5:** vez044.

35 Briant L, Wade CM, Puel J, Brown AJ, Guyader M. Analysis of envelope sequence variants suggests multiple mechanisms of mother-to-child transmission of human immunodeficiency virus type 1. *J Virol* 1995; **69:** 3778–88.

36 Todesco E, Wirden M, Calin R, et al. Caution is needed in interpreting HIV transmission chains by ultradeep sequencing. *AIDS* 2019; **33:** 691–99.

37 Wade CM, Lobidel D, Brown AJ. Analysis of human immunodeficiency virus type 1 env and gag sequence variants derived from a mother and two vertically infected children provides evidence for the transmission of multiple sequence variants. *J Gen Virol* 1998; **79:** 1055–68.

38 Sturdevant CB, Dow A, Jabara CB, et al. Central nervous system compartmentalization of HIV-1 subtype C variants early and late in infection in young children. *PLoS Pathog* 2012; **8:** e1003094–1003094.

39 Rieder P, Joos B, Scherrer AU, et al. Characterization of human immunodeficiency virus type 1 (HIV-1) diversity and tropism in 145 patients with primary HIV-1 infection. *Clin Infect Dis* 2011; **53:** 1271–79.

40 Dukhovlinova E, Masharsky A, Vasileva A, et al. Characterization of the transmitted virus in an ongoing HIV-1 epidemic driven by injecting drug use. *AIDS Res Hum Retroviruses* 2018; **34:** 867–78.

41 Chaillon A, Gianella S, Little SJ, et al. Characterizing the multiplicity of HIV founder variants during sexual transmission among MSM. *Virus Evol* 2016; **2:** vew012.

42 Rossenkhan R, Rolland M, Labuschagne JPL, et al. Combining viral genetics and statistical modeling to improve HIV-1 time-of-infection estimation towards enhanced vaccine efficacy assessment. *Viruses* 2019; **11:** 607.

43 Chen Y, Li N, Zhang T, et al. Comprehensive characterization of the transmitted/founder *env* genes from a single MSM cohort in China. *J Acquir Immune Defic Syndr* 2015; **69:** 403–12.

44 Novitsky V, Moyo S, Wang R, Gaseitsiwe S, Essex M. Deciphering multiplicity of HIV-1C infection: transmission of closely related multiple viral lineages. *PLoS One* 2016; **11:** e0166746.

45 Tovanabutra S, Sirijatuphat R, Pham PT, et al. Deep sequencing reveals central nervous system compartmentalization in multiple transmitted/founder virus acute HIV-1 infection. *Cells* 2019; **8:** 902.

46 Nofemela A, Bandawe G, Thebus R, et al. Defining the human immunodeficiency virus type 1 transmission genetic bottleneck in a region with multiple circulating subtypes and recombinant forms. *Virology* 2011; **415:** 107–13.

47 Herbeck JT, Rolland M, Liu Y, et al. Demographic processes affect HIV-1 evolution in primary infection before the onset of selective processes. *J Virol* 2011; **85:** 7523–34.

48 Smith SA, Burton SL, Kilembe W, et al. Diversification in the HIV-1 envelope hyper-variable domains V2, V4, and V5 and higher probability of transmitted/founder envelope glycosylation favor the development of heterologous neutralization breadth. *PLoS Pathog* 2016; **12:** e1005989–1005989.

49 Poss M, Martin HL, Kreiss JK, et al. Diversity in virus populations from genital secretions and peripheral blood from women recently infected with human immunodeficiency virus type 1. *J Virol* 1995; **69:** 8118–22.

50 Verhofstede C, Demecheleer E, De Cabooter N, et al. Diversity of the human immunodeficiency virus type 1 (HIV-1) *env* sequence after vertical transmission in mother-child pairs infected with HIV-1 subtype A. *J Virol* 2003; **77:** 3050–57.

51 Park SY, Mack WJ, Lee HY. Enhancement of viral escape in HIV-1 Nef by STEP vaccination. *AIDS* 2016; **30:** 2449–58.

52 Danaviah S, de Oliveira T, Bland R, et al. Evidence of long-lived founder virus in mother-to-child HIV transmission. *PLoS One* 2015; **10:** e0120389.

53 Novitsky V, Lagakos S, Herzig M, et al. Evolution of proviral gp120 over the first year of HIV-1 subtype C infection. *Virology* 2009; **383:** 47–59.

54 Long EM, Martin HL Jr, Kreiss JK, et al. Gender differences in HIV-1 diversity at time of infection. *Nat Med* 2000; **6:** 71–75.

55 Salazar-Gonzalez JF, Salazar MG, Keele BF, et al. Genetic identity, biological phenotype, and evolutionary pathways of transmitted/founder viruses in acute and early HIV-1 infection. *J Exp Med* 2009; **206:** 1273–89.

56 Rolland M, Tovanabutra S, deCamp AC, et al. Genetic impact of vaccination on breakthrough HIV-1 sequences from the STEP trial. *Nat Med* 2011; **17:** 366–71.

57 Kishko M, Somasundaran M, Brewster F, Sullivan JL, Clapham PR, Luzuriaga K. Genotypic and functional properties of early infant HIV-1 envelopes. *Retrovirology* 2011; **8:** 67.

58 Deymier MJ, Ende Z, Fenton-May AE, et al. Heterosexual transmission of subtype C HIV-1 selects consensus-like variants without increased replicative capacity or interferon-α resistance. *PLoS Pathog* 2015; **11:** e1005154.

59 Gounder K, Padayachi N, Mann JK, et al. High frequency of transmitted HIV-1 Gag HLA class I-driven immune escape variants but minimal immune selection over the first year of clade C infection. *PLoS One* 2015; **10:** e0119886–0119886.

60 Leda AR, Hunter J, Castro de Oliveira U, et al. HIV-1 genetic diversity and divergence and its correlation with disease progression among antiretroviral naïve recently infected individuals. *Virology* 2020; **541:** 13–24.

61 Kiwelu IE, Novitsky V, Margolin L, et al. HIV-1 subtypes and recombinants in northern Tanzania: distribution of viral quasispecies. *PLoS One* 2012; **7:** e47605.

62 Sagar M, Wu X, Lee S, Overbaugh J. HIV-1 V1-V2 envelope loop sequences expand and add glycosylation sites over the course of infection and these modifications affect antibody neutralization sensitivity. *J Virol* 2006; **80:** 9586–98.

63 Brooks K, Jones BR, Dilernia DA, et al. HIV-1 variants are archived throughout infection and persist in the reservoir. *PLoS Pathog* 2020; **16:** e1008378.

64 Gottlieb GS, Heath L, Nickle DC, et al. HIV-1 variation before seroconversion in men who have sex with men: analysis of acute/early HIV infection in the multicenter AIDS cohort study. *J Infect Dis* 2008; **197:** 1011–15.

65 Delwart E, Magierowska M, Royz M, et al. Homogeneous quasispecies in 16 out of 17 individuals during very early HIV-1 primary infection. *AIDS* 2002; **16:** 189–95.

66 Renjifo B, Chung M, Gilbert P, et al. In-utero transmission of quasispecies among human immunodeficiency virus type 1 genotypes. *Virology* 2003; **307:** 278–82.

67 Wagner GA, Pacold ME, Kosakovsky Pond SL, et al. Incidence and prevalence of intrasubtype HIV-1 dual infection in at-risk men in the United States. *J Infect Dis* 2014; **209:** 1032–38.

68 Sagar M, Lavreys L, Baeten JM, et al. Infection with multiple human immunodeficiency virus type 1 variants is associated with faster disease progression. *J Virol* 2003; **77:** 12921–26.

69 Sterrett S, Learn GH, Edlefsen PT, et al. Low multiplicity of HIV-1 infection and no vaccine enhancement in VAX003 injection drug users. *Open Forum Infect Dis* 2014; **1:** ofu056.

70 Rolland M, Tovanabutra S, Dearlove B, et al. Molecular dating and viral load growth rates suggested that the eclipse phase lasted about a week in HIV-1 infected adults in east Africa and Thailand. *PLoS Pathog* 2020; **16:** e1008179.

71 Baalwa J, Wang S, Parrish NF, et al. Molecular identification, cloning and characterization of transmitted/founder HIV-1 subtype A, D and A/D infectious molecular clones. *Virology* 2013; **436:** 33–48.

72 Ritola K, Pilcher CD, Fiscus SA, et al. Multiple V1/V2 *env* variants are frequently present during primary infection with human immunodeficiency virus type 1. *J Virol* 2004; **78:** 11208–18.

73 Villabona-Arenas ChJ, Hall M, Lythgoe KA, et al. Number of HIV-1 founder variants is determined by the recency of the source partner infection. *Science* 2020; **369**: 103.

74 Dickover RE, Garratty EM, Plaeger S, Bryson YJ. Perinatal transmission of major, minor, and multiple maternal human immunodeficiency virus type 1 variants in utero and intrapartum. *J Virol* 2001; **75:** 2194–203.

75 Sivay MV, Grabowski MK, Zhang Y, et al. Phylogenetic analysis of human immunodeficiency virus from people who inject drugs in Indonesia, Ukraine, and Vietnam: HPTN 074. *Clin Infect Dis* 2020: **71:** 1836–46.

76 Leitner T, Romero-Severson E. Phylogenetic patterns recover known HIV epidemiological relationships and reveal common transmission of multiple variants. *Nat Microbiol* 2018; **3:** 983–88.

77 Kijak GH, Sanders-Buell E, Chenine A-L, et al. Rare HIV-1 transmitted/founder lineages identified by deep viral sequencing contribute to rapid shifts in dominant quasispecies during acute and early infection. *PLoS Pathog* 2017; **13:** e1006510.

78 Iyer SS, Bibollet-Ruche F, Sherrill-Mix S, et al. Resistance to type 1 interferons is a major determinant of HIV-1 transmission fitness. *Proc Natl Acad Sci USA* 2017; **114:** E590–99.

79 Zhang H, Tully DC, Hoffmann FG, He J, Kankasa C, Wood C. Restricted genetic diversity of HIV-1 subtype C envelope glycoprotein from perinatally infected Zambian infants. *PLoS One* 2010; **5:** e9294.

80 Boeras DI, Hraber PT, Hurlston M, et al. Role of donor genital tract HIV-1 diversity in the transmission bottleneck. *Proc Natl Acad Sci USA* 2011; **108:** E1156–63.

81 Wolinsky SM, Wike CM, Korber BT, et al. Selective transmission of human immunodeficiency virus type-1 variants from mothers to infants. *Science* 1992; **255**: 1134–37.

82 Frange P, Meyer L, Jung M, et al. Sexually-transmitted/founder HIV-1 cannot be directly predicted from plasma or PBMC-derived viral quasispecies in the transmitting partner. *PLoS One* 2013; **8:** e69144–69144.

83 deCamp AC, Rolland M, Edlefsen PT, et al. Sieve analysis of breakthrough HIV-1 sequences in HVTN 505 identifies vaccine pressure targeting the CD4 binding site of Env-gp120. *PLoS One* 2017; **12:** e0185959.

84 Collins-Fairclough AM, Charurat M, Nadai Y, et al. Significantly longer envelope V2 loops are characteristic of heterosexually transmitted subtype B HIV-1 in Trinidad. *PLoS One* 2011; **6:** e19995.

85 Love TMT, Park SY, Giorgi EE, Mack WJ, Perelson AS, Lee HY. SPMM: estimating infection duration of multivariant HIV-1 infections. *Bioinformatics* 2016; **32:** 1308–15.

86 Liu Y, Jia L, Su B, et al. The genetic diversity of HIV-1 quasispecies within primary infected individuals. *AIDS Res Hum Retroviruses* 2020; **36:** 440–49.

87 Kwiek JJ, Russell ES, Dang KK, et al. The molecular epidemiology of HIV-1 envelope diversity during HIV-1 subtype C vertical transmission in Malawian mother-infant pairs. *AIDS* 2008; **22:** 863–71.

88 Nowak P, Karlsson AC, Naver L, Bohlin AB, Piasek A, Sönnerborg A. The selection and evolution of viral quasispecies in HIV-1 infected children. *HIV Med* 2002; **3:** 1–11.

89 Rachinger A, Groeneveld PHP, van Assen S, Lemey P, Schuitemaker H. Time-measured phylogenies of *gag, pol* and *env* sequence data reveal the direction and time interval of HIV-1 transmission. *AIDS* 2011; **25:** 1035–39.

90 Oberle CS, Joos B, Rusert P, et al. Tracing HIV-1 transmission: envelope traits of HIV-1 transmitter and recipient pairs. *Retrovirology* 2016; **13:** 62.

91 Novitsky V, Wang R, Margolin L, et al. Transmission of single and multiple viral variants in primary HIV-1 subtype C infection. *PLoS One* 2011; **6:** e16714.

92 Fischer W, Ganusov VV, Giorgi EE, et al. Transmission of single HIV-1 genomes and dynamics of early immune escape revealed by ultra-deep sequencing. *PLoS One* 2010; **5:** e12303.

93 Ashokkumar M, Aralaguppe SG, Tripathy SP, Hanna LE, Neogi U. Unique phenotypic characteristics of recently transmitted HIV-1 subtype C envelope glycoprotein gp120: use of CXCR6 coreceptor by transmitted founder viruses. *J Virol* 2018; **92:** e00063–18.

94 Salazar-Gonzalez JF, Salazar MG, Tully DC, et al. Use of dried blood spots to elucidate full-length transmitted/founder HIV-1 genomes. *Pathog Immun* 2016; **1:** 129–53.

95 Rossenkhan R, Novitsky V, Sebunya TK, Musonda R, Gashe BA, Essex M. Viral diversity and diversification of major non-structural genes *vif, vpr, vpu, tat* exon 1 and *rev* exon 1 during primary HIV-1 subtype C infection. *PLoS One* 2012; **7:** e35491.

96 Learn GH, Muthui D, Brodie SJ, et al. Virus population homogenization following acute human immunodeficiency virus type 1 infection. *J Virol* 2002; **76:** 11953–59.

97 Henn MR, Boutwell CL, Charlebois P, et al. Whole genome deep sequencing of HIV-1 reveals the impact of early minor variants upon immune recognition during acute infection. *PLoS Pathog* 2012; **8:** e1002529.

98 Long EM, Rainwater SMJ, Lavreys L, Mandaliya K, Overbaugh J. HIV type 1 variants transmitted to women in Kenya require the CCR5 coreceptor for entry, regardless of the genetic complexity of the infecting virus. *AIDS Res Hum Retroviruses* 2002; **18:** 567–76.

99 Derdeyn CA, Decker JM, Bibollet-Ruche F, et al. Envelope-constrained neutralization-sensitive HIV-1 after heterosexual transmission. *Science* 2004; **303**: 2019–22.

100 Kearney M, Maldarelli F, Shao W, et al. Human immunodeficiency virus type 1 population genetics and adaptation in newly infected individuals. *J Virol* 2009; **83:** 2715–27.

101 Chaillon A, Gianella S, Wertheim JO, Richman DD, Mehta SR, Smith DM. HIV migration between blood and cerebrospinal fluid or semen over time. *J Infect Dis* 2014; **209:** 1642–52.

102 Zanini F, Brodin J, Thebo L, et al. Population genomics of intrapatient HIV-1 evolution. *eLife* 2015; **4:** e11282.

103 Martinez DR, Tu JJ, Kumar A, et al. Maternal broadly neutralizing antibodies can select for neutralization-resistant, infant-transmitted/founder HIV variants. *MBio* 2020; **11:** e00176–20.

104 Le AQ, Taylor J, Dong W, et al. Differential evolution of a CXCR4-using HIV-1 strain in CCR5wt/wt and CCR5Δ32/Δ32 hosts revealed by longitudinal deep sequencing and phylogenetic reconstruction. *Sci Rep* 2015; **5:** 17607.

105 Balinda SN, Kapaata A, Xu R, et al. Characterization of near full-length transmitted/founder HIV-1 subtype D and A/D recombinant genomes in a heterosexual Ugandan population (2006-2011). *Viruses* 2022; **14:** 334.

106 Thompson RN, Wymant C, Spriggs RA, Raghwani J, Fraser C, Lythgoe KA. Link between the numbers of particles and variants founding new HIV-1 infections depends on the timing of transmission. *Virus Evol* 2019; **5:** vey038.

107 Lee HY, Giorgi EE, Keele BF, et al. Modeling sequence evolution in acute HIV-1 infection. *J Theor Biol* 2009; **261:** 341–60.

108 Slatkin M, Hudson RR. Pairwise comparisons of mitochondrial DNA sequences in stable and exponentially growing populations. *Genetics* 1991; **129:** 555–62.

109 Simon V, Zennou V, Murray D, Huang Y, Ho DD, Bieniasz PD. Natural variation in Vif: differential impact on APOBEC3G/3F and a potential role in HIV-1 diversification. *PLoS Pathog* 2005; **1:** e6.

110 Bourara K, Liegler TJ, Grant RM. Target cell APOBEC3C can induce limited G-to-A mutation in HIV-1. *PLoS Pathog* 2007; **3:** 1477–85.

111 Shaw GM, Hunter E. HIV transmission. *Cold Spring Harb Perspect Med* 2012; **2:** a006965.