HIV co-infection is associated with reduced *Mycobacterium tuberculosis* transmissibility in sub-Saharan Africa

Etthel M. Windels[1,2*], Eddie M. Wampande[3], Moses L. Joloba[3], W. Henry Boom[4], Galo A. Goig[5,6], Helen Cox[7], Jerry Hella[8], Sonia Borrell[5,6], Sebastien Gagneux[5,6☯], Daniela Brites[5,6☯], Tanja Stadler[1,2*☯]

1 Department of Biosystems Science and Engineering, ETH Zürich, Basel, Switzerland
2 Swiss Institute of Bioinformatics, Lausanne, Switzerland
3 Makerere University, Kampala, Uganda
4 Case Western Reserve University and University Hospitals Cleveland Medical Center, Cleveland, USA
5 Swiss Tropical and Public Health Institute, Allschwill, Switzerland
6 University of Basel, Basel, Switzerland
7 University of Cape Town, Cape Town, South Africa
8 Ifakara Health Institute, Dar es Salaam, Tanzania

☯These authors contributed equally as senior authors.
* Corresponding authors. E-mail: etthel.windels@bsse.ethz.ch; tanja.stadler@bsse.ethz.ch

Abstract

Persons living with HIV are known to be at increased risk of developing tuberculosis (TB) disease upon infection with *Mycobacterium tuberculosis* (*Mtb*). However, it has remained unclear how HIV co-infection affects subsequent *Mtb* transmission from these patients. Here, we customized a Bayesian phylodynamic framework to estimate the effects of HIV co-infection on the *Mtb* transmission dynamics from sequence data. We applied our model to four *Mtb* genomic datasets collected in sub-Saharan African countries with a generalized HIV epidemic. Our results confirm that HIV co-infection is a strong risk factor for developing active TB. Additionally, we demonstrate that HIV co-infection is associated with a reduced effective reproductive number for TB. Stratifying the population by CD4+ T-cell count yielded similar results, suggesting that, in this context, CD4+ T-cell count is not a better predictor of *Mtb* transmissibility than HIV infection status. Together, our genome-based analyses complement observational household studies, and firmly establish the negative association between HIV co-infection and *Mtb* transmissibility.

Author summary

Many sub-Saharan African countries have seen a considerable rise in TB incidence since the introduction of HIV, suggesting a strong interaction between HIV and TB epidemics. HIV infection is recognized as an important risk factor for developing TB, but the contribution of HIV-infected TB patients to further *Mtb* transmission is poorly understood. In this study, we analyzed four sets of *Mtb* genomic sequences collected in different countries, including sequences from HIV-negative and HIV-positive TB patients. We applied a phylodynamic model to these sequences, aimed at inferring

transmission dynamics within and between different host populations. While our findings support that HIV is a strong risk factor for TB, we show that HIV-positive TB patients generate a significantly lower number of secondary TB cases than HIV-negative patients. This suggests that HIV-positive patients often act as sinks in *Mtb* transmission chains, while HIV-negative patients are a major source of transmission.

Introduction

The human immunodeficiency virus 1 (HIV) was first introduced into the human population in the beginning of the 20th century through a zoonotic transmission event [1, 2]. Its silent spread in the following decades resulted in a globally established HIV epidemic, disproportionally affecting sub-Saharan Africa [3]. In addition to directly related healthcare challenges, the high prevalence of HIV in these countries has contributed to a strong rise in tuberculosis (TB) incidence rates [4–9]. Accordingly, HIV co-infection in *Mycobacterium tuberculosis* (*Mtb*)-infected patients has been associated with an increased risk of progression to active TB disease, an increased risk of recurrent *Mtb* infection, and an increased TB case-fatality rate [5, 9–13]. Despite our incomplete understanding of the interactions between TB, HIV, and the human immune system, it is widely accepted that the depletion of CD4+ T-cells underlies the high TB susceptibility and mortality in HIV-positive patients [9, 11, 14].

While many studies support this increased susceptibility to TB disease, the effects of HIV co-infection on the generation of secondary TB cases remain poorly understood [9, 15]. The HIV-associated reduction in CD4+ T-cell count has been shown to be associated with an altered TB disease presentation, including lower levels of lung cavitation, lower bacterial loads in the sputum, and a higher likelihood of extrapulmonary TB [9, 16]. This distinct lung pathology could result in reduced *Mtb* transmission, as transmission is mainly driven by the formation of aerosols from infected lungs, and lung cavitations are known to enhance transmission [17, 18].

Several household contact studies have indicated reduced infectiousness of HIV-positive TB index cases [19–24], although notably, many of these studies only considered sputum smear-negative HIV patients or patients with considerably reduced CD4+ T-cell counts [21–24]. In contrast, a meta-analysis [25] and a more recent whole-genome sequence analysis of multidrug-resistant (MDR) *Mtb* isolates [26] found no association between HIV co-infection and the probability of *Mtb* transmission.

While the altered lung pathology of HIV co-infected TB patients could affect the rate at which these patients transmit *Mtb*, the number of secondary cases generated is also determined by how long these patients remain infectious for *Mtb*. Previous studies on the duration of *Mtb* infectiousness in HIV-positive patients showed mixed results. Several studies indicate that HIV co-infection shortens the *Mtb* infectious period [27–29]. This could be explained by a higher TB mortality rate, or faster TB disease progression, resulting in a more timely diagnosis and initiation of treatment [8, 9, 13, 30, 31]. In contrast, one study estimated that the time until accessing TB treatment was longer for HIV-positive patients [32], which could be explained by a postponed diagnosis due to barriers to care for HIV patients, or by higher rates of smear-negativity and an atypical disease presentation.

As the effects of HIV co-infection on the *Mtb* transmission rate and infectious period are not well established, it remains unclear how HIV affects the overall transmissibility of *Mtb*. Here we analyzed four *Mtb* genomic datasets from countries with a high burden of HIV in sub-Saharan Africa (Malawi, South Africa, Tanzania, and Uganda). In particular, we applied a Bayesian phylodynamic model, coupling an epidemiological model with a model of sequence evolution, to investigate how HIV co-infection affects the *Mtb* transmission dynamics. Our phylodynamic model stratifies the TB patient

population by HIV infection status, and is parametrized with the aim of estimating the effect of HIV co-infection on the risk of developing active TB upon exposure, as well as the average number of secondary TB cases generated per patient (i.e., the effective reproductive number). Our results confirm that HIV co-infection is associated with an increased risk of developing active TB, and provides evidence for reduced *Mtb* transmissibility from HIV-positive TB patients.

Results

We analyzed complete *Mtb* genomes collected from TB patients in four sub-Saharan African countries: 1,209 sequences from Karonga District, Malawi (1995-2011) [33], 1,133 sequences from Khayelitsha, Cape Town, South Africa (2008-2018) [34], 1,074 sequences from Temeke District, Dar es Salaam, Tanzania (2013-2019) [35], and 185 sequences from Kampala, Uganda (1995-2012) [36,37] (see Materials and methods for details on the study populations). The sequences from Uganda have been used partially in other studies [38,39] but are analyzed here together for the first time. *Mtb* lineage distributions per sampling location are shown in S1 Table.

To quantify the effects of HIV co-infection on *Mtb* transmission in these locations, we customized a phylodynamic model based on the structured birth-death model [40], with the TB patient population stratified by HIV infection status as determined at the time of TB diagnosis (S1 Fig). In this model, transmission of *Mtb* within and between subpopulations (i.e., HIV-negative and HIV-positive TB patients) is described with different transmission rates (number of transmission events per patient per unit of time), and each subpopulation is additionally characterized by a rate of becoming uninfectious (1/infectious period) and a sampling rate. The ratio of the transmission rate and the becoming uninfectious rate corresponds to the effective reproductive number (R_e), representing the average number of secondary TB cases that one patient generates in the same or the other subpopulation. To explicitly model the effects of HIV, we reparametrized this model with (1) a base R_e, corresponding to the R_e for TB in a purely HIV-negative population (R_e^b), (2) a parameter for the multiplicative effect of HIV co-infection on the R_e of TB patients, at the donor side of transmission (f_1), (3) a parameter for the multiplicative effect of HIV co-infection on the risk of getting diagnosed for active TB disease after *Mtb* exposure (f_2), (4) the rate at which HIV-negative patients become *Mtb* uninfectious (δ_-), (5) a parameter for the multiplicative effect of HIV co-infection on the rate of becoming *Mtb* uninfectious (f_3), and (6) a parameter for the HIV prevalence in the general population (p_{HIV}) (S1 Fig; see Materials and methods for more details on the phylodynamic model). The HIV prevalence was included to account for the HIV-negative and HIV-positive susceptible population sizes, which in turn influence the contact probabilities and thus the R_e in each location. HIV prevalences were not estimated from the genomic data, but set to time-varying levels based on location-specific prevalence data from World Bank [41–44] (S2 Fig). To improve the identifiability of the parameters of interest, the becoming uninfectious rate for HIV-negative patients was fixed to 1 year^{-1}. This corresponds to an average *Mtb* infectious period of 1 year, which is within the range of previous estimates [27–29,32]. We assumed no migration in the model, implying that the HIV status of patients does not change during the course of their *Mtb* infection. We also assumed a constant R_e^b, as justified by the results of an unstructured birth-death skyline analysis [45] (S3 Fig). Each sampling location was analyzed independently. For each location, we inferred *Mtb* lineage-specific phylogenetic trees, with each tree modelled as having an independent origin and evolutionary parameters. Epidemiological parameters were assumed to be the same for all lineages co-circulating within a given location. All parameters were estimated with the Bayesian phylogenetics package BEAST2 [46,47],

with prior distributions summarized in S2 Table.

The posterior maximum clade credibility trees of *Mtb* isolates from Tanzania are displayed in Fig 1 (trees for the other locations are shown in S4 Fig, S5 Fig, and S6 Fig), indicating limited clustering of HIV-positive patients. The posterior distributions of the parameter estimates show that the TB disease development risk for HIV-positive patients relative to HIV-negative patients is significantly higher than 1 in each sampled location, with posterior means ranging from 4.48 to 11.49 (Fig 2a). These results suggest that HIV-positive patients have a 4- to 11-fold increased risk of developing active TB upon exposure, which is in accordance with the increased TB incidence rate observed in HIV patients [9, 10, 12, 13]. Posterior estimates for the relative R_e (HIV-positive relative to HIV-negative patients) are all significantly lower than 1, with posterior means ranging from 0.047 to 0.20, implying that HIV co-infection is associated with a 5- to 21-fold reduction in R_e for TB (Fig 2b). This could result from an altered TB disease presentation, as supported by the significantly lower chest X-ray scores of HIV-positive compared to HIV-negative TB patients in Uganda (Welch's two-sample t-test, $p = 0.044$). Similarly, the HIV-positive patients from Tanzania showed reduced chest X-ray scores (Welch's two-sample t-test, $p < 0.001$), less cavity development (χ^2-test, $p < 0.001$), and lower bacterial loads in the sputum (Welch's two-sample t-test, $p = 0.0027$) [35], reflecting a distinct lung pathology, consistent with a reduced infectiousness. Furthermore, clinical data from Malawi and South Africa showed that HIV-positive patients were strongly associated with extrapulmonary TB (χ^2-test, $p < 0.001$ for both datasets), which is non-transmissible. All these observations are consistent with previous studies [9, 16].

Fig 1. Posterior maximum clade credibility trees of *Mtb* isolates from Tanzania. Posterior maximum clade credibility trees per lineage, summarizing the posterior tree distribution resulting from the phylodynamic analyses on the *Mtb* sequences from Tanzania, with tips labeled by HIV infection status. Trees of isolates from the other locations are shown in S4 Fig, S5 Fig, and S6 Fig.

Fig 2. Phylodynamic estimates of the effects of HIV co-infection on *Mtb* transmission. Prior (grey) and posterior (coloured) distributions per sampling location of the estimates for a) the relative risk of developing active TB upon exposure (HIV-positive patients relative to HIV-negative patients), b) the relative R_e for TB. For all posterior distributions, the 95% HPD intervals do not contain 1.

We could not identify a clear effect of HIV co-infection on the *Mtb* infectious period, due to wide posterior distributions for some locations and conflicting results for others (S7 Fig). However, irrespective of the infectious period, HIV co-infection was associated with a reduced *Mtb* transmission rate (S7 Fig), suggesting that HIV mainly affects the strength, rather than the period of infectiousness.

To investigate how much our results are impacted by the assumptions on the becoming uninfectious rate of HIV-negative patients, we repeated the analyses with fixed values of 0.5 year^{-1} and 2 year^{-1} for this parameter (S8 Fig). Further sensitivity analyses comprised different priors on the parameters for the HIV effects on *Mtb* transmission and TB disease development risk (S9 Fig), and fixing the clock rate to 10^{-8} resp. 10^{-7} substitutions per site per year (S10 Fig; see S3 Table for the clock rate estimates resulting from the main analyses). While the absolute values of the posterior estimates of interest were weakly dependent on the choice of priors, all sensitivity analyses resulted in the same qualitative conclusions regarding the relative R_e and relative progression risk.

Phylodynamic birth-death estimates are not only informed by the genomic data, but also by the distribution of sample collection dates. An additional set of analyses where the sequences were ignored showed that estimates of the relative progression risk with and without the genomic data were in close agreement, with only slightly shifted posterior distributions and more certainty in the estimates when the sequences were included (S11 Fig), suggesting that they contain little information about this parameter. For the relative R_e, including the sequences resulted in shifted and narrower posterior distributions (S11 Fig), indicating that while the isolation dates and HIV infection status of patients are the major source of information, the genomic data further inform this parameter.

To identify potential biases introduced through model assumptions and priors, we repeated the analyses on datasets where the HIV status of the patients was permuted. These datasets still contained signal for the relative progression risk and relative R_e (S12 Fig). This can be explained by the HIV prevalence being 4 to 7 times higher in the sampled TB patients than in the general population in the countries under study. Indeed, randomly assigning the HIV status using the average HIV frequency in the general population during the sampling period resulted in posterior distributions for the relative progression risk and relative R_e that overlap with 1, implying no effect of HIV co-infection (S12 Fig). Together, these results demonstrate that the signal for the HIV effect parameters originates from the data rather than from model assumptions, with the parameter inference presumably being driven by 1) the high prevalence of HIV within the population of TB patients, and 2) the sampling dates and sequences informing the overall R_e estimate, which in turn constrains the HIV effect parameters.

HIV patients might show different levels of CD4+ T-cell depletion, depending on the stage of HIV infection and whether the patient is on antiretroviral therapy (ART). Several studies have indicated that decreased CD4+ T-cell counts are associated with a reduced frequency of lung cavitations (see [9] for an overview), and we found a similar association in HIV patients from Uganda (Welch's two-sample t-test, $p = 0.030$), suggesting that CD4+ T-cell counts might be a better predictor for *Mtb* transmissibility and TB progression than the HIV infection status. In contrast, one study showed that TB incidence rates were increased even in HIV patients with high CD4+ T-cell counts [11], suggesting that other aspects of HIV infection might also play a role. To investigate the contribution of CD4+ T-cell counts to the observed effects of HIV co-infection on *Mtb* transmission, we repeated our phylodynamic analyses on sequences from South Africa and Uganda, with subpopulations defined by CD4+ T-cell counts (lower resp. higher than 350 cells/μl, the threshold recommended by WHO to prioritize patients for ART) instead of HIV status (S13 Fig). For the other sampling locations, CD4+ T-cell counts were not available. These analyses resulted in similar posterior means and HPD intervals for the parameters of interest (Fig 3), suggesting that the CD4+ T-cell count can be used as a predictor of *Mtb* transmission, but that, within the context of our analyses, it is not more informative than the HIV status.

Fig 3. CD4+ T-cell count as predictor for *Mtb* transmissibility and TB disease progression. Prior (grey) and posterior (coloured) distributions per sampling location of the estimates for a) the relative risk of developing active TB upon exposure (patients with low CD4+ T-cell count relative to patients with high CD4+ T-cell count), b) the relative R_e for TB. A threshold of 350 CD4+ T-cells/μl was used to classify patients. For all posterior distributions, the 95% HPD intervals do not contain 1.

Discussion

The effects of HIV co-infection on *Mtb* transmission have remained elusive, with previous studies yielding contradictory results. Here we used a phylodynamic approach to address the question based on *Mtb* sequences and HIV/*Mtb* co-infection data sampled in four different African countries. Our phylodynamic analyses confirm that HIV/*Mtb* co-infected individuals are at high risk of developing active TB disease compared to HIV-negative *Mtb*-infected individuals. Moreover, we found that HIV-positive TB patients on average cause significantly fewer secondary TB cases compared to HIV-negative TB patients. These findings were reproduced across all four countries.

Our finding that HIV co-infection is a strong risk factor for developing active TB disease upon exposure explains why many TB epidemics in sub-Saharan Africa seem to be driven by the high HIV prevalence in these settings [4–9]. The underlying cause of this increased susceptibility to TB might be the depletion of CD4+ T-cells in HIV patients [9, 11, 14]. As we could only investigate the overall risk of developing active TB disease after contact with a TB patient, it remains unclear, based on our data, whether HIV also affects the risk of *Mtb* infection.

The consistently reduced TB R_e from HIV-positive patients, observed in all countries under study, seems to be linked to a reduced number of transmission events per patient per unit of time. These findings are in accordance with the reduced *Mtb* infectiousness of HIV patients previously observed in various household contact studies [19–24]. A reduced infectiousness of HIV-positive TB patients can potentially be explained by an altered TB disease presentation in HIV-positive patients, which could in turn result from an impaired immune system. This notion is supported by significant associations between HIV infection status and clinical variables related to lung damage and bacterial burden, observed in this and previous studies [9, 16].

The TB R_e of HIV-positive TB patients might be additionally reduced through a shorter infectious period, due to more rapid disease progression and/or an increased mortality rate [8, 9, 13, 30, 31]. In contrast, increased bacterial drug resistance, delayed diagnosis due to an atypical disease presentation, and barriers to care for HIV patients could increase the *Mtb* infectious period of HIV-positive TB patients [32]. While we could not identify a consistent HIV effect on the infectious period, we showed that the reduced R_e of HIV-positive TB patients was linked to a lower transmission rate, irrespective of the duration of the infectious period.

The effects of HIV co-infection might be complicated by ART, which alleviates the CD4+ T-cell depletion in HIV patients [48, 49]. As information on ART was lacking for most patients, we could not directly take this into account in our analyses. However, we assumed that the CD4+ T-cell measurements from the patients in South Africa and Uganda would reflect differences in ART. As the CD4+ T-cell count classification (low/high) for these patients largely overlapped with their HIV infection status (S13 Fig), we could not identify any distinguishable effect of CD4+ T-cell counts. ART only recently became widely accessible to HIV patients in South Africa and Uganda [50, 51], which might explain why CD4+ T-cell count and HIV status are largely redundant. In other settings, ART might play a more important role in determining CD4+ T-cell counts, and consequently, CD4+ T-cell counts might be a better predictor of *Mtb* transmissibility.

HIV co-infection has been shown to be associated with rifampicin-resistant TB, potentially due to increased resistance acquisition during TB treatment [52]. While the prevalence of drug resistance during the sampling period was low in Malawi, Tanzania, and Uganda, the South-African dataset consists of rifampicin-resistant *Mtb* isolates only, indicating that HIV is associated with reduced *Mtb* transmissibility irrespective of drug resistance. Other potential confounders are poverty-related risk factors, patient sex, and patient age. However, with our current approach, these confounders are challenging to

control for due to the rapidly increasing model complexity.

Notably, the major source of information on HIV effects in the model are the sampling dates and the HIV status of the TB patients, indicating that our customized birth-death model would have been able to capture the signal in the data even in the absence of *Mtb* sequences. A potential explanation for this limited signal in the sequence data is the fact that HIV status is a host-related factor that is not associated to the bacterial genetic background. Consequently, the *Mtb* genomes of HIV-positive TB patients are dispersed across the phylogenetic tree, resulting in high uncertainty on ancestral states and thus few informative branching events.

One limitation of our approach is that we did not account for multiple *Mtb* introductions into the study populations, nor for changes in the TB R_e over time. However, HIV was most likely introduced only after the establishment of different *Mtb* lineages, and the majority of branching events informing the epidemiological parameters occur after the introduction of HIV, suggesting minimal impact on the estimates. In support of this notion, current evidence indicates that the main *Mtb* lineages circulating in these parts of Africa were introduced several centuries ago [35, 38, 53–56]. Moreover, no biases were observed when randomly re-assigning the HIV status of patients. A second limitation is that our estimates might be biased due to an underestimation of the HIV prevalence in the general population. As these potential biases originate from the input data rather than the model, they cannot be identified with our randomization approach. A third limitation is the assumption that the probabilities of contact between and within HIV-positive and HIV-negative subpopulations are solely a function of the size of each subpopulation, ignoring any preferential contacts due to social effects. Finally, a fourth limitation of the model is the assumption that patients infected with *Mtb* immediately become infectious (i.e., no period of latent infection). While this assumption could affect the interpretation of the transmission rate and infectious period, the relative R_e is expected to be robust, even if the duration of the latent period would be associated with HIV infection. Notably, our model does not distinguish between *Mtb*-uninfected patients and patients who are infected but never develop active TB disease.

Taken together, our results demonstrate that a high HIV prevalence can fuel a TB epidemic by increasing the risk for TB disease progression in HIV-positive patients, but that these patients do not proportionally contribute to further *Mtb* transmission. HIV-positive patients can thus be considered as 'sinks' in transmission chains. By contrast, HIV-negative TB patients serve as the 'sources' by being disproportionally responsible for *Mtb* transmission. Our findings have implications for TB control, and call for a particular attention to HIV-negative TB patients, ideally through active case finding, thereby ensuring that these patients are diagnosed and treated as early as possible to prevent further spread of the disease.

Materials and methods

Study populations

All datasets in this study consist of whole genome sequences (Illumina) of *Mtb* strains collected in countries with a generalized HIV epidemic.

Malawi Raw Illumina reads were retrieved from the European Nucleotide Archive (project accession numbers ERP000436 and ERP001072). These sequences were obtained from adults with culture-confirmed TB diagnosed at the hospital and peripheral health centres in Karonga District, northern Malawi between 1995 and 2011 [33]. Information about the HIV status of the patients was kindly provided by the

authors of the study, and only sequences from patients with known HIV status were retained ($n = 1,209$). The incidence of smear-positive TB in adults in the district during the sampling period corresponds to 87-124 cases per 100,000 people per year [33].

South Africa We used previously sequenced isolates from a retrospective cohort study of individuals routinely diagnosed with rifampicin-resistant (RR) or multidrug-resistant (MDR) TB in Khayelitsha, Cape Town, South Africa between 2008 and 2018 (raw reads available in the European Nucleotide Archive under project accession numbers PRJEB45389 and PRJNA670836) [34]. Only sequences from patients with known HIV status were retained ($n = 1,133$). The TB notification rate in Khayelitsha was estimated around 80 RR/MDR cases per 100,000 people per year [52].

Tanzania We used previously sequenced isolates from a cohort of sputum smear-positive and GeneXpert-positive adult TB patients prospectively recruited at the Temeke District hospital in Dar es Salaam, Tanzania between 2013 and 2019 (raw reads available in the European Nucleotide Archive under project accession number PRJEB49562) [35]. Only sequences from patients with known HIV status were retained ($n = 1,074$). The TB notification rate in Temeke in 2020 was 3,994 cases per year (Jerry Hella, personal communication).

Uganda Bacterial isolates and clinical data were obtained from TB patients recruited in two large household contact studies. An initial study was conducted from 1995 to 1999 to describe the epidemiology of TB in urban Kampala, Uganda [36,57]. The second is known as the "Kawempe Community Health" study which ran from 2000 to 2012 [37]. We whole-genome sequenced 185 isolates belonging to *Mtb* sublineage 4.6.1 (raw reads available in the European Nucleotide Archive under project accession numbers PRJEB11460, PRJNA354716, and PRJEB64921). These isolates were obtained from HIV-positive and HIV-negative TB patients defined mostly as index cases within a household (with the exception of 4 and 3 strains from contact and co-prevalent cases, respectively). HIV status was determined by ELISA and confirmed by Western blot at baseline. CD4+ T-cell counts were available for 48 patients. The incidence rate of sputum smear-positive TB in Kampala in 2001-2002 was estimated around 370 cases per 100,000 people per year [36]. As only the year of sample isolation was available, all isolates were assumed to be collected on the 1st of January of the corresponding year. The institutional review boards at University Hospitals of Cleveland in the United States and AIDS Research Council in Uganda reviewed the study protocols and final approval was obtained from the Uganda National Council for Science and Technology. Written informed consent was obtained from all patients that participated in the study. All participants were given appropriate pre- and post-test HIV counseling and AIDS education. The protocols and the procedures for the protection of human subjects were approved by the Uganda National Council Ethics Committee and the Institutional Ethics Review Board at Makerere University, Kampala.

Whole-genome sequencing

Bacterial isolates from Kampala, Uganda were cultured on Middlebrook 7H10 supplemented with 10% glycerol and OADC until confluent colonies appeared on the plates. The colonies were scraped off and their genomic DNA was extracted using the CTAB method [58]. Selected strains were whole-genome sequenced on an Illumina HiSeq 2000 instrument at the commercial facility GATC (Germany). Library preparation was performed according to Illumina's TruSeq DNA Sample Preparation Guide (Illumina, San Diego, CA). Single-end sequence reads of approximately 50 bp

were obtained. Demultiplexing was performed automatically by the CASAVA pipeline v1.8.0 (Illumina, San Diego, CA).

WGS analyses and alignments

For all datasets, the Illumina reads were processed and analyzed as described in [34, 35]. Lineages and sublineages were identified using the SNP-based classification by Steiner et al. [59]. For all sequences per lineage and per location, an alignment of polymorphic positions was assembled by concatenating all high-quality SNPs. Sites that had more than 10% of missing data, as well as drug-resistance-related sites, were excluded from the alignment.

Phylodynamic analyses

We fit a multitype birth-death model to the sequence alignments [40], with two types corresponding to HIV-negative TB patients and HIV-positive TB patients (S1 Fig). Under this model, a 'birth' event corresponds to an *Mtb* transmission event from one host to another, which can occur within and between types. A 'death' event occurs when a host becomes uninfectious for *Mtb* due to recovery or death. The model was parametrized with the R_e within a purely HIV-negative population (R_e^b), the rate at which HIV-negative patients become uninfectious (δ_-), the multiplicative effect of HIV co-infection on transmitting *Mtb* (f_1), the multiplicative effect of HIV co-infection on the risk of TB disease development upon exposure (f_2), and the multiplicative effect of HIV co-infection on the rate of becoming uninfectious for *Mtb* (f_3). The effective reproductive number for TB within the HIV-negative subpopulation, within the HIV-positive subpopulation, from the HIV-negative to the HIV-positive subpopulation, and from the HIV-positive to the HIV-negative subpopulation (R_e^{--}, R_e^{++}, R_e^{-+}, and R_e^{+-}, respectively), as well as the rate at which HIV-positive patients become *Mtb* uninfectious (δ_+), are then as follows:

$$\begin{aligned} R_e^{--} &= (1 - p_{\text{HIV}})R_e^b \\ R_e^{++} &= f_1 f_2 p_{\text{HIV}} R_e^b \\ R_e^{-+} &= f_2 p_{\text{HIV}} R_e^b \\ R_e^{+-} &= f_1(1 - p_{\text{HIV}})R_e^b \\ \delta_+ &= f_3 \delta_- \end{aligned}$$

p_{HIV} represents the overall prevalence of HIV in the general population (including both *Mtb*-infected and *Mtb*-uninfected individuals) and is included to account for different sizes of the HIV-negative and HIV-positive populations. As the HIV prevalence in a country changed over time since the date of HIV introduction, we let p_{HIV} change at three different time points in the past, according to HIV prevalence data from World Bank [41–44] (S2 Fig). Hence, the effective reproductive numbers also changed through time.

The overall reproductive numbers for HIV-negative and HIV-positive patients are stated below. From these equations, it can be seen that $f_1 = R_e^+/R_e^-$.

$$\begin{aligned} R_e^- &= R_e^{--} + R_e^{-+} = (1 - p_{\text{HIV}})R_e^b + f_2 p_{\text{HIV}} R_e^b \\ R_e^+ &= R_e^{+-} + R_e^{++} = f_1(1 - p_{\text{HIV}})R_e^b + f_1 f_2 p_{\text{HIV}} R_e^b \end{aligned}$$

Transmission rates (i.e., rates of *Mtb* transmission per patient per unit of time) from HIV-negative and HIV-positive TB patients can be obtained using the definition of the reproductive number:

$$\lambda_- = R_e^- \delta_-$$
$$\lambda_+ = R_e^+ \delta_+$$

We assumed no migration between subpopulations, implying that HIV-negative patients cannot get infected with HIV during their period of *Mtb* infectiousness. TB patients are sampled with sampling proportion s, which was set equal to zero before the onset of sampling. Upon sampling an infected patient, the patient is assumed to become uninfectious with probability r [60].

We further assumed a strict molecular clock and a general time-reversible nucleotide substitution model with four gamma rate categories to account for site-to-site rate heterogeneity (GTR+Γ_4).

We performed phylodynamic inference using the bdmm package [40] in BEAST v2.6.6 [46, 47]. Data from each location were analyzed independently. For each location, variable SNP alignments were generated per *Mtb* lineage and augmented with a count of invariant A, C, G, and T nucleotides to avoid ascertainment bias [61]. To avoid unreasonably long runtimes, any alignment containing more than 400 sequences was randomly downsampled to 400 sequences, and sampling proportion priors were adjusted accordingly. Population dynamic parameters were inferred jointly for the different *Mtb* lineages within one location: each lineage was represented with an independent tree with its own origin time and nucleotide substitution parameters, but sharing all other parameters with the other lineages.

Three independent Markov Chain Monte Carlo chains were run for each analysis, with states sampled every 1,000 steps. Tracer [62] was used to assess convergence and confirm that the effective sample size (ESS) was at least 200 for the parameters of interest. 10% of each chain was discarded as burn-in, and the remaining samples across the three chains were pooled using LogCombiner [47], resulting in at least 300,000,000 iterations in combined chains.

Prior distributions

All parameters and their corresponding prior distributions are listed in S2 Table. For the sampling proportion, a uniform prior was chosen with lower bound set to zero and upper bound set equal to the ratio of the number of sequences, corrected for downsampling, and the total number of reported cases during the sampling period (S4 Table).

Sensitivity analyses

The robustness of the phylodynamic inference was assessed by changing the fixed value of δ_- to 0.5 and 2 year^{-1}, by changing the prior on f_1 and f_2 to a Lognormal(0,0.5) distribution, and by fixing the clock rate to 10^{-8} and 10^{-7} substitutions per site per year. To evaluate the relative impact of the sequence data on our parameter estimates, a phylodynamic analysis was performed using the same setup as the main analyses, but without any sequence data.

Birth-death skyline analysis

To investigate whether R_e^b can be assumed constant through time, we ran a birth-death skyline analysis on sequences from the most abundant lineage per sampling location [45]. No population structure is assumed in this model. Two time intervals were used to estimate potential changes in the overall R_e over time, with the change point set at the estimated time of HIV introduction (S2 Fig). The overall rate of becoming *Mtb* uninfectious was assumed constant through time. For the sampling proportion, clock

model and substitution model parameters, the same settings and priors were used as in the multitype birth-death model.

Randomization of HIV status

The HIV infection status of the patients was randomized in two ways. First, the HIV status labels were permuted, implying that the HIV prevalence among the sampled TB patients was kept unchanged. Second, the HIV status was randomly assigned to each patient, with an overall HIV prevalence among the patients assumed equal to the average HIV prevalence in the general population (including both *Mtb*-uninfected and *Mtb*-infected individuals) during the sampling period (S2 Fig) [41–44]. Each of the randomization procedures was replicated 10 times.

Population stratification by CD4+ T-cell count

The TB patient population was stratified based on CD4+ T-cell count, with a threshold set at 350 cells/μl, corresponding to the threshold recommended by WHO to prioritize patients for ART (S13 Fig) [63]. As CD4+ T-cell counts were not monitored for HIV-negative TB patients from Uganda, these patients were all classified as having a high CD4+ T-cell count (in accordance with data from South Africa, Figure). The fitted phylodynamic model was equivalent to the model based on HIV status, with HIV-negative patients being replaced by patients with a high CD4+ T-cell count (≥ 350 cells/μl) and HIV-positive patients being replaced by patients with a low CD4+ T-cell count (< 350 cells/μl). Correspondingly, p_{HIV} was replaced by $p_{\text{lowCD4+}}$, the prevalence of patients with low CD4+ T-cell counts. This prevalence was estimated as follows: the HIV prevalence in the general population was multiplied by the observed proportion of patients with low CD4+ T-cell count among HIV-positive patients in the dataset. $p_{\text{lowCD4+}}$ was set to 75% of this value, as HIV-positive patients with low CD4+ T-cell counts are likely overrepresented among TB patients. Changing this 75% to higher or lower values did not change the qualitative conclusions.

Statistical analyses

Associations between HIV infection status and other variables were tested using Welch's t-tests and χ^2-tests implemented in R.

Data availability

185 genome sequences collected in Kampala, Uganda, were deposited to the European Nucleotide Archive (ENA) at EBI, registered under project accession numbers PRJEB11460 (`https://www.ebi.ac.uk/ena/browser/view/PRJEB11460`), PRJNA354716 (`https://www.ebi.ac.uk/ena/browser/view/PRJNA354716`) and PRJEB64921 (`https://www.ebi.ac.uk/ena/browser/view/PRJEB64921`).

Code availability

The code for the phylodynamic analyses, including BEAST2 XML files, is available at `https://github.com/EtthelWindels/tb_hiv`.

Acknowledgments

E.M.W. and T.S. received funding from ETH Zürich and from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme grant agreement No. 101001077 (PhyCogy). This study was further supported by grants from ERC (883582) and the Swiss National Science Foundation (310030_188888) both to S.G., and a Swiss and South Africa joint research award to H.C. and S.G. (IZLSZ3_170834). Calculations were performed on the Euler cluster at ETH Zürich and at sciCORE (`http://scicore.unibas.ch/`) scientific computing core facility at the University of Basel. We would like to thank Louis du Plessis for valuable feedback on the manuscript, Timothy G. Vaughan for help with the phylodynamic analyses, and the Malawi Epidemiology and Intervention Research Unit for providing patient data.

References

1. Worobey M, Gemmel M, Teuwen DE, Haselkorn T, Kunstman K, Bunce M, et al. Direct evidence of extensive diversity of HIV-1 in Kinshasa by 1960. Nature. 2008;455(7213):661–664.

2. Faria NR, Rambaut A, Suchard MA, Baele G, Bedford T, Ward MJ, et al. The early spread and epidemic ignition of HIV-1 in human populations. Science. 2014;346(6205):56–61.

3. UNAIDS. Global AIDS update 2023. Geneva: UNAIDS; 2023.

4. Corbett EL, Watt CJ, Walker N, Maher D, Williams BG, Raviglione MC, et al. The growing burden of tuberculosis. Archives of Internal Medicine. 2003;163(9):1009.

5. Corbett EL, Marston B, Churchyard GJ, De Cock KM. Tuberculosis in sub-Saharan Africa: Opportunities, challenges, and change in the era of antiretroviral treatment. The Lancet. 2006;367(9514):926–937.

6. Karim SSA, Churchyard GJ, Karim QA, Lawn SD. HIV infection and tuberculosis in South Africa: an urgent need to escalate the public health response. The Lancet. 2009;374(9693):921–933.

7. Bekker LG, Wood R. The changing natural history of tuberculosis and HIV coinfection in an urban area of hyperendemicity. Clinical Infectious Diseases. 2010;50(S3):S208–S214.

8. Getahun H, Gunneberg C, Granich R, Nunn P. HIV infection-associated tuberculosis: The epidemiology and the response. Clinical Infectious Diseases. 2010;50(S3):S201–S207.

9. Kwan CK, Ernst JD. HIV and tuberculosis: A deadly human syndemic. Clinical Microbiology Reviews. 2011;24(2):351–376.

10. World Health Organization. Global tuberculosis report 2009. Geneva: World Health Organization; 2019.

11. Lawn SD, Myer L, Edwards D, Bekker LG, Wood R. Short-term and long-term risk of tuberculosis associated with CD4 cell recovery during antiretroviral therapy in South Africa. AIDS. 2009;23(April):1717–1725.

12. UNAIDS. Global AIDS update 2020. Geneva: UNAIDS; 2020.

13. World Health Organization. Global tuberculosis report 2022. Geneva: World Health Organization; 2022.

14. Diedrich CR, Flynn JL. HIV-1/*Mycobacterium tuberculosis* coinfection immunology: How does HIV-1 exacerbate tuberculosis? Infection and Immunity. 2011;79(4):1407–1417.

15. Peters JS, Andrews JR, Hatherill M, Hermans S, Martinez L, Schurr E, et al. Advances in the understanding of *Mycobacterium tuberculosis* transmission in HIV-endemic settings. The Lancet Infectious Diseases. 2019;19(3):e65–e76.

16. Sterling TR, Pham PA, Chaisson RE. HIV infection-related tuberculosis: Clinical manifestations and treatment. Clinical Infectious Diseases. 2010;50:223–230.

17. Jones-López EC, Namugga O, Mumbowa F, Ssebidandi M, Mbabazi O, Moine S, et al. Cough aerosols of *Mycobacterium tuberculosis* predict new infection: A household contact study. American Journal of Respiratory and Critical Care Medicine. 2013;187(9):1007–1015.

18. Jones-López EC, Acuña-Villaorduña C, Ssebidandi M, Gaeddert M, Kubiak RW, Ayakaka I, et al. Cough aerosols of *Mycobacterium tuberculosis* in the prediction of incident tuberculosis disease in household contacts. Clinical Infectious Diseases. 2016;63:10–20.

19. Espinal MA, Peréz EN, Baéz J, Hénriquez L, Fernández K, Lopez M, et al. Infectiousness of *Mycobacterium tuberculosis* in HIV-1-infected patients with tuberculosis: A prospective study. Lancet. 2000;355(9200):275–280.

20. Carvalho ACC, DeRiemer K, Nunes ZB, Martins M, Comelli M, Marinoni A, et al. Transmission of *Mycobacterium tuberculosis* to contacts of HIV-infected tuberculosis patients. American Journal of Respiratory and Critical Care Medicine. 2001;164(12):2166–2171.

21. Kenyon TA, Creek T, Laserson K, Makhoa M, Chimidza N, Mwasekaga M, et al. Risk factors for transmission of *Mycobacterium tuberculosis* from HIV-infected tuberculosis patients, Botswana. International Journal of Tuberculosis and Lung Disease. 2002;6(10):843–850.

22. Huang CC, Tchetgen ET, Becerra MC, Cohen T, Hughes KC, Zhang Z, et al. The effect of HIV-related immunosuppression on the risk of tuberculosis transmission to household contacts. Clinical Infectious Diseases. 2014;58(6):765–774.

23. Martinez L, Sekandi JN, Castellanos ME, Zalwango S, Whalen CC. Infectiousness of HIV-seropositive patients with tuberculosis in a high-burden African setting. American Journal of Respiratory and Critical Care Medicine. 2016;194(9):1152–1163.

24. Martinez L, Woldu H, Chen C, Hallowell BD, Castellanos ME, Lu P, et al. Transmission dynamics in tuberculosis patients with human immunodeficiency virus: A systematic review and meta-analysis of 32 observational studies. Clinical Infectious Diseases. 2021;73(9):e3446–55.

25. Cruciani M, Malena M, Bosco O, Giorgio G, Serpelloni G. The impact of human immunodeficiency virus type 1 on infectiousness of tuberculosis: A meta-analysis. Clinical Infectious Diseases. 2001;33(11):1922–1930.

26. Eldholm V, Rieux A, Monteserin J, Lopez JM, Palmero D, Lopez B, et al. Impact of HIV co-infection on the evolution and transmission of multidrug-resistant tuberculosis. eLife. 2016;5:e16644.

27. Corbett EL, Charalambous S, Moloi VM, Fielding K, Grant AD, Dye C, et al. Human immunodeficiency virus and the prevalence of undiagnosed tuberculosis in African gold miners. American Journal of Respiratory and Critical Care Medicine. 2004;170(6):673–679.

28. Corbett EL, Bandason T, Yin BC, Munyati S, Godfrey-Faussett P, Hayes R, et al. Epidemiology of tuberculosis in a high HIV prevalence population provided with enhanced diagnosis of symptomatic disease. PLoS Medicine. 2007;4(1):e22.

29. Corbett EL, Bandason T, Cheung YB, Makamure B, Dauya E, Munyati SS, et al. Prevalent infectious tuberculosis in Harare, Zimbabwe: Burden, risk factors and implications for control. International Journal of Tuberculosis and Lung Disease. 2009;13(10):1231–1237.

30. Ackah AN, Coulibaly D, Digbeu H, Diallo K, Vetter KM, Coulibaly IM, et al. Response to treatment, mortality, and CD4 lymphocyte counts in HIV-infected persons with tuberculosis in Abidjan, Cote d'Ivoire. The Lancet. 1995;345:607–610.

31. DeRiemer K, Kawamura LM, Hopewell PC, Daley CL. Quantitative impact of human immunodeficiency virus infection on tuberculosis dynamics. American Journal of Respiratory and Critical Care Medicine. 2007;176:936–944.

32. Wood R, Middelkoop K, Myer L, Grant AD, Whitelaw A, Lawn SD, et al. Undiagnosed tuberculosis in a community with high HIV prevalence: Implications for tuberculosis control. American Journal of Respiratory and Critical Care Medicine. 2007;175(1):87–93.

33. Guerra-Assunção J, Crampin A, Houben R, Mzembe T, Mallard K, Coll F, et al. Large-scale whole genome sequencing of *M. tuberculosis* provides insights into transmission in a high prevalence area. eLife. 2015;2015(4):e05166.

34. Goig GA, Menardo F, Salaam-Dreyer Z, Dippenaar A, Streicher EM, Daniels J, et al. Effect of compensatory evolution in the emergence and transmission of rifampicin-resistant *Mycobacterium tuberculosis* in Cape Town, South Africa: a genomic epidemiology study. The Lancet Microbe. 2023;5247(23):1–10.

35. Zwyer M, Rutaihwa LK, Windels E, Hella J, Menardo F, Sasamalo M, et al. Back-to-Africa introductions of *Mycobacterium tuberculosis* as the main cause of tuberculosis in Dar es Salaam, Tanzania. PLoS Pathogens. 2023;19(4):e1010893.

36. Guwatudde D, Zalwango S, Kamya MR, Debanne SM, Diaz MI, Okwera A, et al. Burden of tuberculosis in Kampala, Uganda. Bulletin of the World Health Organization. 2003;81(11):799–805.

37. Wampande EM, Mupere E, Debanne SM, Asiimwe BB, Nsereko M, Mayanja H, et al. Long-term dominance of *Mycobacterium tuberculosis* Uganda family in peri-urban Kampala-Uganda is not associated with cavitary disease. BMC Infectious Diseases. 2013;13(1):484.

38. Stucki D, Brites D, Jeljeli L, Coscolla M, Liu Q, Trauner A, et al. *Mycobacterium tuberculosis* lineage 4 comprises globally distributed and geographically restricted sublineages. Nature Genetics. 2016;48(12):1535–1543.

39. Kabahita JM, Kabugo J, Kakooza F, Adam I, Guido O, Byabajungu H, et al. First report of whole-genome analysis of an extensively drug-resistant *Mycobacterium tuberculosis* clinical isolate with bedaquiline, linezolid and clofazimine resistance from Uganda. Antimicrobial Resistance and Infection Control. 2022;11(1):1–8.

40. Kühnert D, Stadler T, Vaughan TG, Drummond AJ. Phylodynamics with migration: A computational framework to quantify population structure from genomic data. Molecular Biology and Evolution. 2016;33(8):2102–2116.

41. World Bank. Prevalence of HIV, total (% of population ages 15-49) - Malawi; 2022. Available from: `https://data.worldbank.org/indicator/SH.DYN.AIDS.ZS?locations=MW`.

42. World Bank. Prevalence of HIV, total (% of population ages 15-49) - South Africa; 2022. Available from: `https://data.worldbank.org/indicator/SH.DYN.AIDS.ZS?locations=ZA`.

43. World Bank. Prevalence of HIV, total (% of population ages 15-49) - Tanzania; 2022. Available from: `https://data.worldbank.org/indicator/SH.DYN.AIDS.ZS?locations=TZ`.

44. World Bank. Prevalence of HIV, total (% of population ages 15-49) - Uganda; 2022. Available from: `https://data.worldbank.org/indicator/SH.DYN.AIDS.ZS?locations=UG`.

45. Stadler T, Kühnert D, Bonhoeffer S, Drummond AJ. Birth-death skyline plot reveals temporal changes of epidemic spread in HIV and hepatitis C virus (HCV). Proceedings of the National Academy of Sciences of the United States of America. 2013;110(1):228–233.

46. Bouckaert R, Heled J, Kühnert D, Vaughan T, Wu CH, Xie D, et al. BEAST 2: A software platform for Bayesian evolutionary analysis. PLoS Computational Biology. 2014;10(4):e1003537.

47. Bouckaert R, Vaughan TG, Barido-Sottani J, Duchêne S, Fourment M, Gavryushkina A, et al. BEAST 2.5: An advanced software platform for Bayesian evolutionary analysis. PLoS Computational Biology. 2019;15(4):1–28.

48. Autran B, Carcelain G, Li TS, Blanc C, Mathez D, Tubiana R, et al. Positive effects of combined antiretroviral therapy on CD4+ T cell homeostasis and function in advanced HIV disease. Science. 1997;277(5322):112–116.

49. Carcelain G, Debré P, Autran B. Reconstitution of CD4+ T lymphocytes in HIV-infected individuals following antiretroviral therapy. Current Opinion in Immunology. 2001;13(4):483–488.

50. World Bank. Antiretroviral therapy coverage (South Africa; 2023. Available from: `https://data.worldbank.org/indicator/SH.HIV.ARTC.ZS?locations=ZA`.

51. World Bank. Antiretroviral therapy coverage (Uganda; 2023. Available from: `https://data.worldbank.org/indicator/SH.HIV.ARTC.ZS?locations=UG`.

52. Cox H, Salaam-Dreyer Z, Goig GA, Nicol MP, Menardo F, Dippenaar A, et al. Potential contribution of HIV during first-line tuberculosis treatment to subsequent rifampicin-monoresistant tuberculosis and acquired tuberculosis drug resistance in South Africa: a retrospective molecular epidemiology study. The Lancet Microbe. 2021;5247(21):1–10.

53. Brynildsrud OB, Pepperell CS, Suffys P, Grandjean L, Monteserin J, Debech N, et al. Global expansion of *Mycobacterium tuberculosis* lineage 4 shaped by colonial migration and local adaptation. Science Advances. 2018;4(10):1–12.

54. O'Neill MB, Shockey A, Zarley A, Aylward W, Eldholm V, Kitchen A, et al. Lineage specific histories of *Mycobacterium tuberculosis* dispersal in Africa and Eurasia. Molecular Ecology. 2019;28(13):3241–3256.

55. Rutaihwa LK, Menardo F, Stucki D, Gygli SM, Ley SD, Malla B, et al. Multiple introductions of *Mycobacterium tuberculosis* Lineage 2-Beijing into Africa over centuries. Frontiers in Ecology and Evolution. 2019;7(MAR).

56. Menardo F, Gagneux S, Rutaihwa LK, Zwyer M, Borrell S, Comas I, et al. Local adaptation in populations of *Mycobacterium tuberculosis* endemic to the Indian Ocean Rim. F1000Research. 2021;10:1–24.

57. Guwatudde D, Nakakeeto M, Jones-Lopez EC, Maganda A, Chiunda A, Mugerwa RD, et al. Tuberculosis in household contacts of infectious cases in Kampala, Uganda. American Journal of Epidemiology. 2003;158(9):887–898.

58. Ausubel M, Brent R, Kingston R, Moore D, Seidman J, Smith J, et al. Current Protocols in Molecular Biology. New York: Greene Publishing Associated and Wiley Interscience; 1987.

59. Steiner A, Stucki D, Coscolla M, Borrell S, Gagneux S. KvarQ: Targeted and direct variant calling from fastq reads of bacterial genomes. BMC Genomics. 2014;15:881.

60. Gavryushkina A, Welch D, Stadler T, Drummond AJ. Bayesian inference of sampled ancestor trees for epidemiology and fossil calibration. PLoS Computational Biology. 2014;10(12):e1003919.

61. Leaché AD, Banbury BL, Felsenstein J, De Oca ANM, Stamatakis A. Short tree, long tree, right tree, wrong tree: New acquisition bias corrections for inferring SNP phylogenies. Systematic Biology. 2015;64(6):1032–1047.

62. Rambaut A, Drummond AJ, Xie D, Baele G, Suchard MA. Posterior summarization in Bayesian phylogenetics using Tracer 1.7. Systematic Biology. 2018;67(5):901–904.

63. World Health Organization. Consolidated guidelines on the use of antiretroviral drugs for treating and preventing HIV infection 2016. Geneva: World Health Organization; 2015.

Supporting information

S1 Fig. Schematic representation of the phylodynamic model. Phylodynamic model used to estimate HIV effects on *Mtb* transmission, based on a structured birth-death model with HIV-negative and HIV-positive TB patients representing different subpopulations. Each subpopulation has its own rate of becoming uninfectious (indicated as δ_- and δ_+) and sampling rate (indicated as s_- and s_+). Transmission events occur within each subpopulation with reproductive numbers indicated as R_e^{--} and R_e^{++}, and between subpopulations with effective reproductive numbers indicated as R_e^{-+} and R_e^{+-}. b) For the analyses in this study, the model was reparametrized by expressing the reproductive numbers as a function of a base R_e (R_e^b),

the HIV prevalence in the general population (p_{HIV}), the multiplicative effect of HIV co-infection on the R_e of TB patients (f_1), the multiplicative effect of HIV co-infection on the risk of developing active TB when exposed (f_2), and the multiplicative effect of HIV co-infection on the rate of becoming uninfectious (f_3).

S2 Fig. HIV prevalence in different countries over time. Coloured lines represent the prevalence per country over time, as reported by World Bank [41–44], while the dashed grey lines represent the values used in our phylodynamic model.

S3 Fig. R_e estimates resulting from the birth-death skyline analyses. Prior (grey) and posterior (coloured) distributions per sampling location of the estimates of the overall R_e before and after the estimated time of HIV introduction into the country, assuming no structure in the population.

S4 Fig. Posterior maximum clade credibility trees of *Mtb* isolates from Malawi. Posterior maximum clade credibility tree per lineage, summarizing the posterior tree distribution resulting from the phylodynamic analyses on the sequences from Malawi, with tips labeled by HIV infection status.

S5 Fig. Posterior maximum clade credibility trees of *Mtb* isolates from South Africa. Posterior maximum clade credibility tree per lineage, summarizing the posterior tree distribution resulting from the phylodynamic analyses on the sequences from South Africa, with tips labeled by HIV infection status.

S6 Fig. Posterior maximum clade credibility tree of *Mtb* isolates from Uganda. Posterior maximum clade credibility tree, summarizing the posterior tree distribution resulting from the phylodynamic analyses on the sequences from Uganda (lineage 4 only), with tips labeled by HIV infection status.

S7 Fig. Additional parameter estimates resulting from the main phylodynamic analyses. Posterior distributions per sampling location of the estimates for a) the relative *Mtb* infectious period (HIV-positive relative to HIV-negative patients), b) the relative *Mtb* transmission rate. For all posterior distributions in (b), the 95% HPD intervals do not contain 1.

S8 Fig. Parameter estimates for the sensitivity analyses on the becoming uninfectious rate. Prior (grey) and posterior (coloured) distributions per sampling location of the estimates for a) the relative risk of developing active TB upon exposure (HIV-positive relative to HIV-negative patients), assuming a fixed becoming uninfectious rate of 0.5 year^{-1}, b) the relative R_e for TB, assuming a fixed becoming uninfectious rate of 0.5 year^{-1}, c) the relative risk of developing active TB upon exposure, assuming a fixed becoming uninfectious rate of 2 year^{-1}, d) the relative R_e for TB, assuming a fixed becoming uninfectious rate of 2 year^{-1}. For all posterior distributions, the 95% HPD intervals do not contain 1.

S9 Fig. Parameter estimates for the sensitivity analyses on the HIV effect priors. Prior (grey) and posterior (coloured) distributions per sampling location of the estimates for a) the relative risk of developing active TB upon exposure (HIV-positive relative to HIV-negative patients), and b) the relative R_e for TB, assuming Lognormal(0,0.5) priors on the effect of HIV on *Mtb* transmission (f_1) and TB disease progression (f_2). For all posterior distributions, the 95% HPD intervals do not contain 1.

S10 Fig. Parameter estimates for the sensitivity analyses on the clock rate. Prior (grey) and posterior (coloured) distributions per sampling location of the estimates for a) the relative risk of developing active TB upon exposure (HIV-positive relative to HIV-negative patients), assuming a fixed clock rate of 10^{-8} substitutions per site per year, b) the relative R_e for TB, assuming a fixed clock rate of 10^{-8} substitutions per site per year, c) the relative risk of developing active TB upon exposure, assuming a fixed clock rate of 10^{-7} substitutions per site per year, d) the relative R_e for TB, assuming a fixed clock rate of 10^{-7} substitutions per site per year. For all posterior distributions, the 95% HPD intervals do not contain 1.

S11 Fig. Parameter estimates for the analyses with and without genomic data. a) Posterior distributions per sampling location of the estimates for the relative risk of developing active TB upon exposure (HIV-positive relative to HIV-negative patients), only based on the sampling dates and HIV infection status (light colours), or also including the sequences (dark colours). b) Posterior distributions per sampling location of the estimates for the relative R_e for TB, only based on the sampling dates and HIV infection status (light colours), or also including the sequences (dark colours).

S12 Fig. Parameter estimates for the analyses on randomized datasets. Prior (grey) and posterior (coloured) distributions per sampling location of the estimates for a) the relative risk of developing active TB upon exposure (HIV-positive relative to HIV-negative patients), on 10 different datasets where the HIV status labels of the patients were permuted, b) the relative R_e for TB, on 10 different datasets where the HIV status labels of the patients were permuted, c) the relative risk of developing active TB upon exposure, on 10 different datasets where the HIV status labels were randomly assigned using the average HIV frequency in the general population during the sampling period, d) the relative R_e for TB, on 10 different datasets where the HIV status labels were randomly assigned using the average HIV frequency in the general population during the sampling period.

S13 Fig. CD4+ T-cell counts of HIV-negative and HIV-positive TB patients from South Africa and Uganda. In Uganda, CD4+ T-cell counts were only recorded for HIV-positive patients. The dashed line represents the threshold (350 cells/μl) recommended by WHO to prioritize patients for ART [63]. This threshold was used to stratify the TB patient population.

S1 Table. Observed lineage distribution at the different sampling locations, based on the number of sequences in the datasets.

S2 Table. Prior distributions for the parameters of the multitype birth-death model.

S3 Table. Clock rate estimates resulting from the main phylodynamic analyses.

S4 Table. Total number of reported cases during the sampling period and total number of sequences included in the analyses at the different sampling locations.

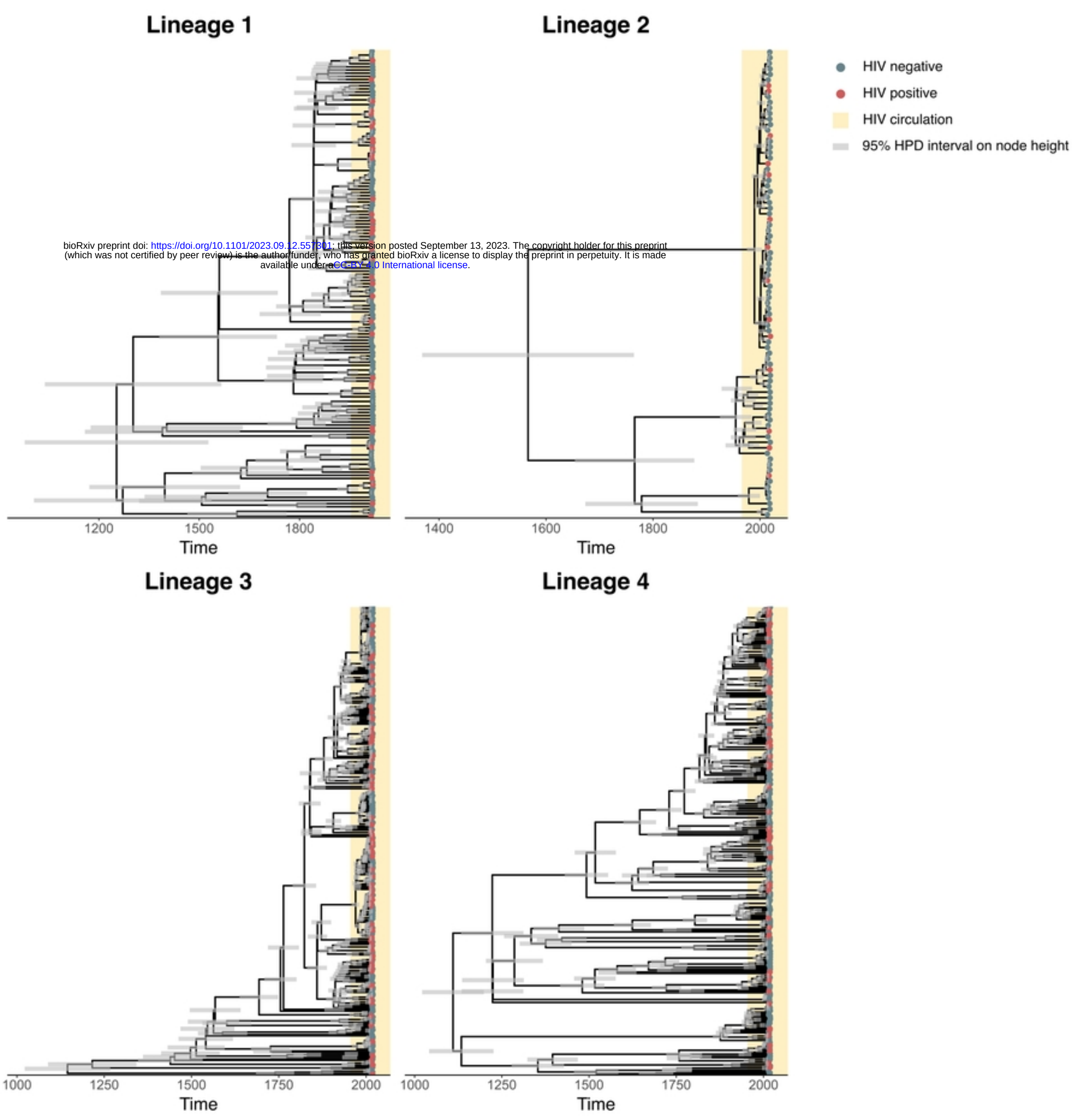

Figure 1

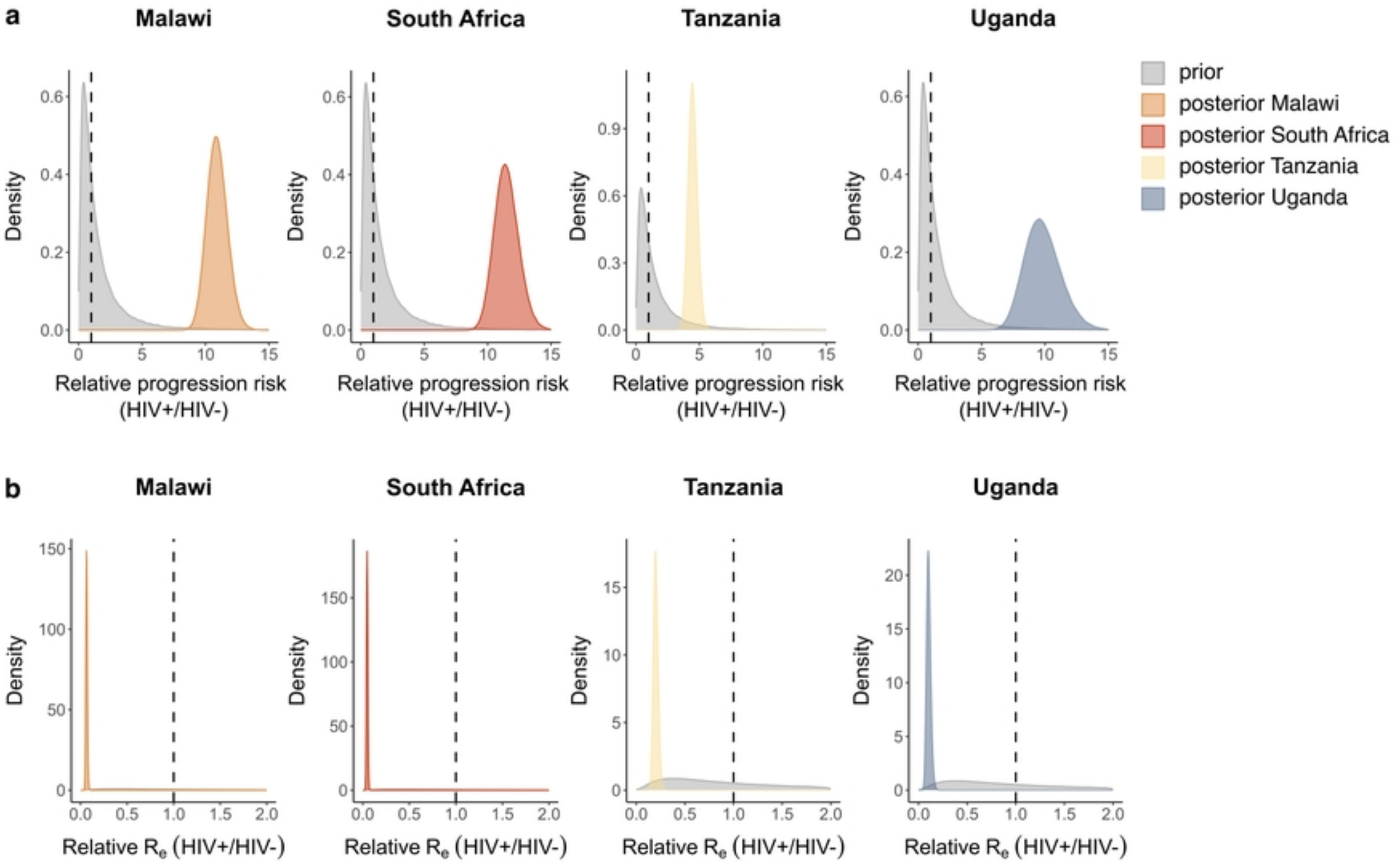
a
Malawi
South Africa
Tanzania
Uganda
prior
posterior Malawi
posterior South Africa
posterior Tanzania
posterior Uganda
Density
Relative progression risk
(HIV+/HIV-)
b
Relative R_e (HIV+/HIV-)

Figure 2

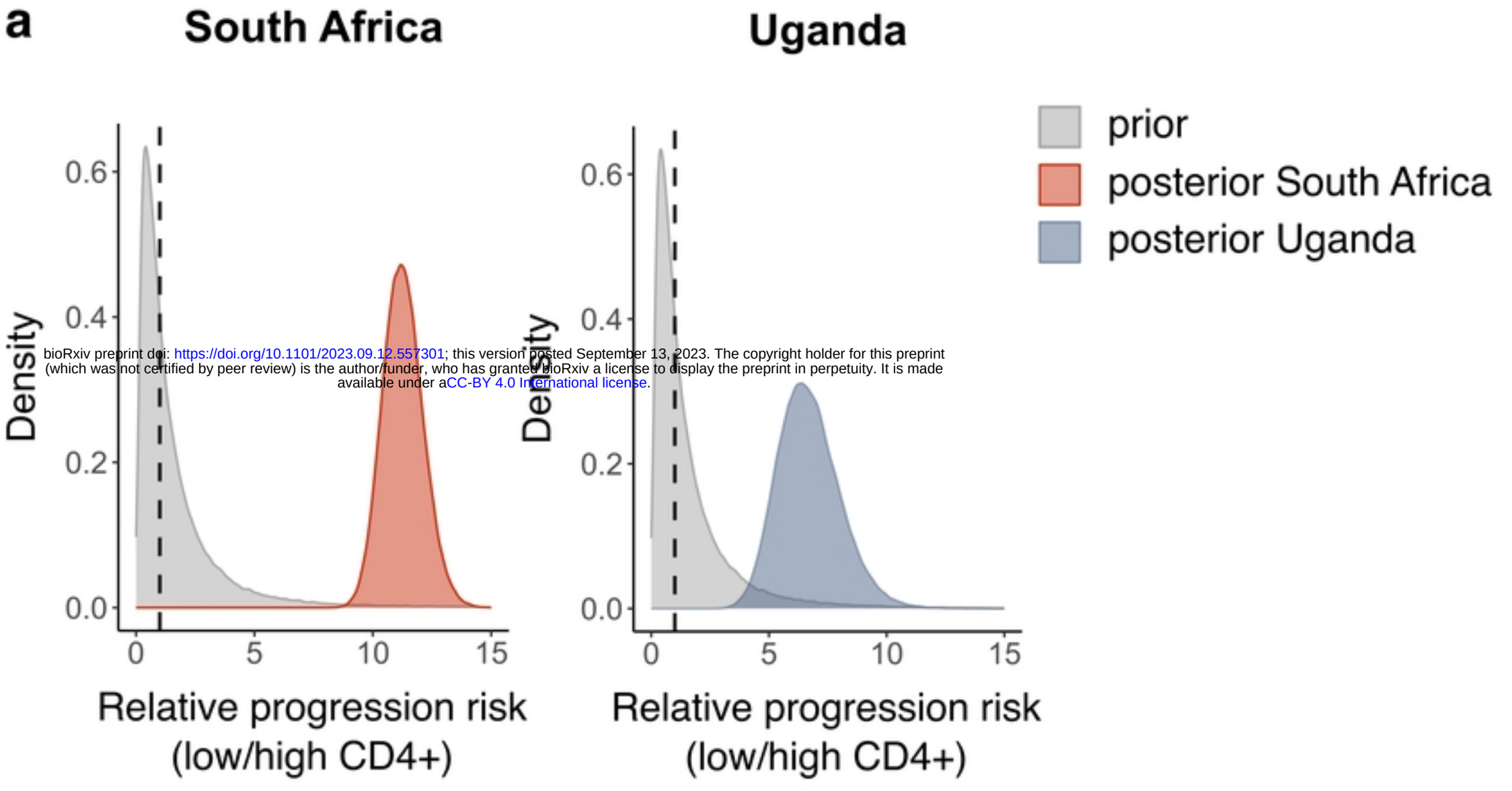

Figure 3