A Practical Guide to Estimating the Heritability of Pathogen Traits

Venelin Mitov*,[1,2] and Tanja Stadler[1,2]

[1]Department of Biosystems Science and Engineering, ETH Zürich, Basel, Switzerland
[2]Swiss Institute of Bioinformatics, Switzerland

*Corresponding author: E-mail: vmitov@gmail.com.
Associate editor: Sergei Kosakovsky Pond

Abstract

Pathogen traits, such as the virulence of an infection, can vary significantly between patients. A major challenge is to measure the extent to which genetic differences between infecting strains explain the observed variation of the trait. This is quantified by the trait's broad-sense heritability, H^2. A recent discrepancy between estimates of the heritability of HIV-virulence has opened a debate on the estimators' accuracy. Here, we show that the discrepancy originates from model limitations and important lifecycle differences between sexually reproducing organisms and transmittable pathogens. In particular, current quantitative genetics methods, such as donor–recipient regression of surveyed serodiscordant couples and the phylogenetic mixed model (PMM), are prone to underestimate H^2, because they neglect or do not fit to the loss of resemblance between transmission partners caused by within-host evolution. In a phylogenetic analysis of 8,483 HIV patients from the United Kingdom, we show that the phenotypic correlation between transmission partners decays with the amount of within-host evolution of the virus. We reproduce this pattern in toy-model simulations and show that a phylogenetic Ornstein–Uhlenbeck model (POUMM) outperforms the PMM in capturing this correlation pattern and in quantifying H^2. In particular, we show that POUMM outperforms PMM even in simulations without selection—as it captures the mentioned correlation pattern—which has not been appreciated until now. By cross-validating the POUMM estimates with ANOVA on closest phylogenetic pairs, we obtain $H^2 \approx 0.2$, meaning ∼20% of the variation in HIV-virulence is explained by the virus genome both for European and African data.

Key words: HIV, set-point viral load (spVL), donor–recipient regression, ANOVA, phylogenetic mixed model, Ornstein–Uhlenbeck.

Introduction

Pathogens transmitted between donor and recipient hosts are genetically related much like children are related to their parents through inherited genes. This analogy between transmission and biological reproduction has inspired the use of heritability (H^2)—a term borrowed from quantitative genetics (Falconer and Mackay 1996; Lynch and Walsh 1998; Hartl and Clark 2007) to measure the contribution of pathogen genetic factors to pathogen traits, such as virulence, transmissibility, and drug-resistance of infections.

Two families of methods have been used to estimate the heritability of a pathogen trait in the absence of knowledge about its genetic basis:

- Resemblance estimators measuring the relative trait-similarity within groups of transmission-related patients. Common methods of that kind are linear regression of donor–recipient (DR) couples (Fraser et al. 2014; Leventhal and Bonhoeffer 2016) and analysis of variance (ANOVA) of patients linked by (near-)identity of carried strains (Anderson et al. 2010; Shirreff et al. 2013).
- Phylogenetic comparative methods measuring the so called phylogenetic heritability, that is, the association

between observed trait values from patients and their (approximate) transmission tree inferred from pathogen sequences. Common examples of such methods are the Felsenstein's independent contrasts (Felsenstein 1985), the phylogenetic mixed model (PMM) (Housworth et al. 2004), and the Pagel's λ (Freckleton et al. 2002).

Most of these methods have been applied in studies of the viral contribution to virulence in an HIV infection (Tang et al. 2004; Alizon et al. 2010; Hecht et al. 2010; Hollingsworth et al. 2010; van der Kuyl et al. 2010; Lingappa et al. 2013; Shirreff et al. 2013; Yue et al. 2013; Fraser et al. 2014; Hodcroft et al. 2014; Bonhoeffer et al. 2015; Leventhal and Bonhoeffer 2016; Bachmann et al. 2017; Bertels et al. 2018; Blanquart et al. 2017). To quantify the virulence of an HIV infection, the above studies have used measurements of the $\lg_{10}$ set point viral load, lg(spVL)—the amount of virions per blood-volume stabilizing in HIV patients at the beginning of the asymptomatic phase and best-predicting its duration (Mellors et al. 1996). In the view of discrepant reports of lg(spVL)-heritability, many authors have questioned the accuracy of the existing methods and have proposed various adaptations of these methods in order to overcome potential pitfalls, such as false model assumptions (e.g., neutral evolution and ultrametricity

of transmission trees) and imperfections in the data (e.g., small data size, presence of cofactors, and measurement error) (Shirreff et al. 2013; Fraser et al. 2014; Hodcroft et al. 2014; Leventhal and Bonhoeffer 2016; Mitov and Stadler 2016; Bachmann et al. 2017; Bertels et al. 2018; Blanquart et al. 2017). Despite these efforts, to date, there is no consensus about the root cause of the discrepancy in lg(spVL)-heritability estimates and there is little reuse of the tools previously implemented, making it hard to compare the estimates from different studies.

In the remainder of the introduction, we consider the definition of broad-sense heritability from the point of view of the key differences between sexually reproducing organisms and clonally transmitted pathogens. Then, in New Approaches, we introduce a model of an epidemic that allows exploring how one of these differences—the within-host evolution of pathogens—affects most of the currently used estimators of heritability. In Results, we compare these estimators based on simulations of the above model and report an analysis of spVL data from a large HIV cohort. In the light of these results, we designate the most reliable estimators of pathogen trait heritability and establish a lower bound for the viral genetic contribution to set-point viral load.

Differences between Pathogens and Sexual Species When Estimating Heritability

According to quantitative genetics theory, the *broad-sense heritability*, H^2, of a quantitative trait is defined in the context of a population of organisms as the ratio of the genotypic over phenotypic variance:

$$H^2 = \mathrm{Var}(G)/\mathrm{Var}(z), \qquad (1)$$

where z denotes the phenotypic value and G denotes the genotypic value assigned to each individual in the population (Falconer and Mackay 1996; Lynch and Walsh 1998; Hartl and Clark 2007). In the case of epidemics, the population represents a sample of hosts, that is, organisms infected by a given type of pathogen. The phenotypic value, z, represents a numerical trait resulting from the infection, and the genotypic value, G, is defined for each pathogen genotype (strain), as the phenotypic value to be expected if it would be measured in a randomly chosen host infected with this strain.

In a large enough population with fully known pathogen genotypes, H^2 could be measured by the direct heritability estimator—the coefficient of determination, R^2_{adj}, obtained over a grouping of the population by genotype. In practice though, this is impossible, because population sizes are small compared with large numbers of (usually unknown) genotypes. To tackle this problem, pathogenecists have relied on the apparent analogy between parent–offspring couples in sexually reproducing populations and DR couples in infected populations. This analogy has motivated the use of correlation measures, such as the DR regression slope, b, and the intraclass correlation in phylogenetic pairs, r_A, to estimate the heritability of pathogen traits (Anderson et al. 2010; Shirreff

et al. 2013; Fraser et al. 2014; Leventhal and Bonhoeffer 2016). However, three differences between the lifecycles of clonally transmitted pathogens and sexually reproducing organisms challenge this approach:

Asexual Haploid Nature of Pathogen Transmission
The first difference is that, unlike the reproduction of diploid organisms, the transmission of a pathogen from a donor to a recipient is more similar to asexual (haploid) reproduction, because, typically, whole pathogens get transferred between hosts.

Partial Quasispecies Transmission
The second difference is that the transmitted proportion of genetic information characterizing the pathogen in the donor is unknown and varying between transmission events. For example, for slowly evolving bacteria such as *Mycobacterium tubercolosis* (Mtb), transmission can be clonal (Bjorn-Mortensen et al. 2016), whereas, for rapidly evolving retroviruses like HIV, transmission is often accompanied by bottlenecks causing only a tiny sample of the large and genetically diverse virus population in the donor (a.k.a., quasispecies) to penetrate and survive in the recipient (Keele et al. 2008).

Within-Host Pathogen Evolution
The third difference involves the change in phenotypic value due to within-host pathogen mutation and recombination. Although genetic change is rare during the lifetime of animals and plants and its phenotypic effects are typically delayed to the offspring generations, it constitutes a hallmark in the lifecycle of pathogens and causes a gradual or immediate phenotypic change such as increasing virulence, immune escape, or drug resistance (fig. 1).

The net outcome of these differences is that unlike family members, for which the amount of genetic overlap is a known constant, for example, 50% for a parent–child couple, the genetic overlap between the two quasispecies in a DR couple is an unknown variable. If there were full quasispecies transmission and no within-host evolution, the pathogen populations found in a donor and a recipient at any moment after a transmission event would be similar to identical twins raised in separate environments. By analogy with twins, any measure of the trait correlation in transmission couples, such as b and r_A, would estimate the broad-sense heritability, H^2 (Lynch and Walsh 1998). However, the partial quasispecies transmission and the within-host evolution taking place in the time between transmission and measurement can lead to a change in the correlation between couple members without affecting H^2 at the population level. We presume that this issue has been at the origin of the discrepancy in previous reports of lg(spVL)-heritability. In particular, the applied methods vary substantially in how they account for the within-host evolution taking place between transmission and measurement: some of them neglect it (Shirreff et al. 2013; Leventhal and Bonhoeffer 2016); others diminish its effect through preferential sampling of patients in the early phase of infection
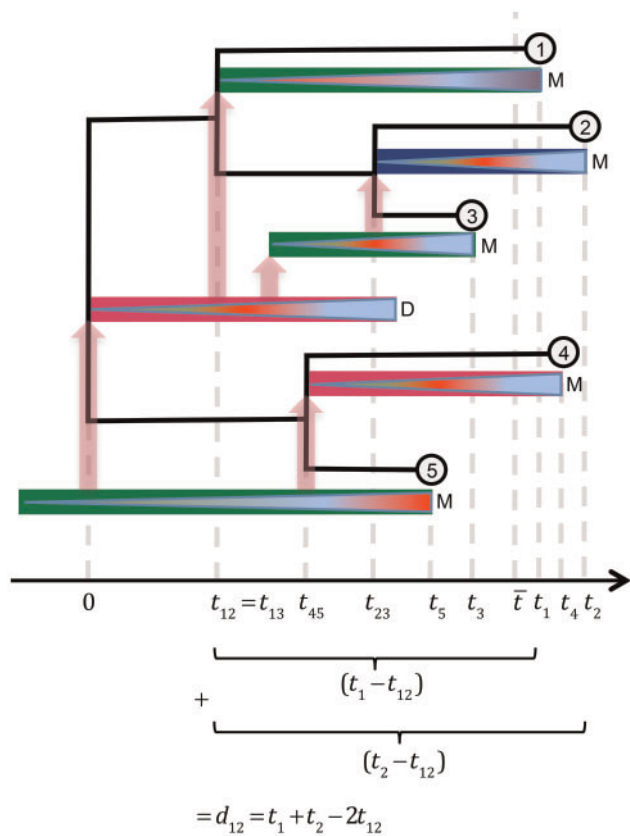
Fig. 1. A schematic representation of an epidemic. Colored rectangles represent infectious periods of hosts, different colors corresponding to different host types. Triangles inside hosts represent pathogen quasispecies, change of color indicating substitution of dominant strains. Capital letters denote host-events: M: diagnosis followed by immediate phenotype measurement, treatment and quarantine for the host; D: host death. The transmission tree connecting the measured hosts is drawn in black. Notice that, due to incomplete sampling, there is no one-to-one correspondance between transmission events and branching points on the tree. By convention, the time origin is at the root of the tree and the time is assumed to increase toward the the tips. We denote by t_i the time distance from the root to tip i. The mean root-tip distance is denoted by $\bar{t}$. For each couple of tips, i and j, we denote by t_{ij} the time distance from the root to their most recent common ancestor (mrca) and by d_{ij} their phylogenetic distance. For clarity, we show how d_{ij} can be expressed in terms of t_{ij} and the root-tip times, t_i and t_j. Couples of tips that are each other's closest tip by phylogenetic distance, for example, (2, 3) and (4, 5), are called "phylogenetic pairs" (PPs). In balanced trees, PPs tend to coincide with pairs of tips descending from the same parent node (a.k.a., siblings or "cherries").

or transmission couples shortly after seroconversion (Hecht et al. 2010; Hollingsworth et al. 2010); third ones attempt to account for it by taking advantage of stochastic models of trait evolution, such as Brownian motion (BM) (Alizon et al. 2010; Hodcroft et al. 2014) or Ornstein–Uhlenbeck (OU) (Mitov and Stadler 2016; Bertels et al. 2018; Blanquart et al. 2017). In the next section, we introduce a simulation based method allowing for within-host evolution, which enables comparing these methods against the direct heritability estimator, R^2_{adj}.

New Approaches

A Toy-Model of an Epidemic

We propose a simulation based method for evaluating different heritability estimators. Our approach differs substantially from previous simulation studies, where the pathogen genotype is equivalent to the genotypic value, G, and is modeled by a continuous branching stochastic process evolving along a given transmission tree (Alizon et al. 2010; Shirreff et al. 2013; Hodcroft et al. 2014; Leventhal and Bonhoeffer 2016). In contrast, we implement a more explicit model in which the pathogen genotype represents a randomly mutating sequence of gene variants (alleles) and the trait value results from the interaction between the pathogen genotype and the host. The main advantages of this approach are 1) the possibility to compare different estimates of H^2 to its true value obtained from the direct estimator, R^2_{adj}, and 2) the possibility to study the effect of within-host mutation and measurement delay on all estimates. As a limitation, the proposed model omits coexistence of strains within a host and partial quasispecies transmission, because of their complexity and the current lack of empirical knowledge and data (see Discussion). For this reason and because of its minimalistic design, we refer to this model as a "toy-model."

In the toy-model, we think of an infection as an asexually reproducing haploid organism. The environment for this organism is the infected host, and the reproduction represents the clonal transmission of the infecting strain to other susceptible hosts. The pathogen has a genome composed of a finite number of loci, which mutate sporadically during infection, resulting in mutant strains. Depending on the within-host fitness of a mutant, it can be eliminated or it can immediately substitute the strain currently invading the host. A trait, z, is determined by the additive effects and epistatic interactions between the alleles at the loci in the genome as well as the interaction between these alleles and the host immune system. The immune system represents a combination of an immutable host type interacting in a predefined way with each possible strain and a randomly drawn host-specific effect, summarizing the unknown effects of other host-related factors, such as age, sex, and habitat. We assume two equally frequent host-types and two trait-determining loci in the pathogen genotype with $M_1=3$ and $M_2=2$ possible alleles at each locus. Thus, there are six possible strains and a total of 12 host type×strain combinations (fig. 2A).

The dynamics of the model combine within-host events, such as strain mutation and substitution, and between-host events, such as transmission, natural, and pathogen-induced death as well as diagnosis followed by immediate uninfectiousness, recovery, and immunity for the patient. These events are modeled as Poisson processes for every infected individual (fig. 2B). The between-host dynamics are inspired from a classical Susceptible-Infected-Recovered (SIR) model with finite population size (ch. 1 in Keeling and Rohani 2007). The main difference with this epidemiological model is that the rate of transmission and the expected infectious period for an infected host can depend on the current trait value and
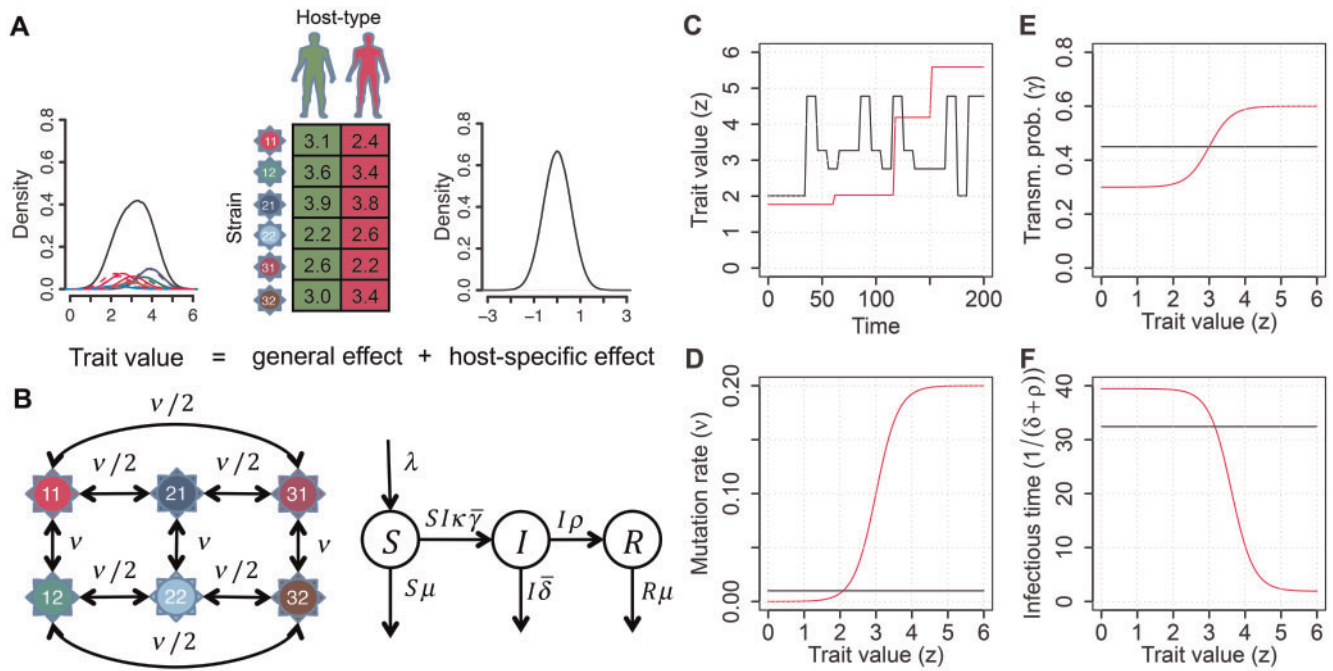
FIG. 2. A toy model of an epidemic with within-host mutation and SIR dynamics. (*A*) A pathogen trait represents the sum of a general <host type×strain> effect and a normally distributed host-specific effect. Pathogen strains are denoted by the alleles at the two loci, for example, "31" stays for allele 3 at locus 1 and allele 1 at locus 2. The density of the trait in a population of hosts represents a mixture of normal densities corresponding to the host type×strain combinations scaled by their relative frequencies. (*B*) Within a host (left), each locus of the infecting strain mutates at a rate ν; horizontal or curved arrows denote mutations at the first locus, vertical arrows denote mutations at the second locus; the rates above the arrows correspond to the per locus mutation rate (ν) divided by the number of possible other alleles at the locus. At the between-host level (right), the alive population is divided into a Susceptible, Infected, and Recovered compartments, letters S, I and R denoting the corresponding proportions in the population at a given time. New individuals become susceptible at a constant rate λ; risky contacts occur at rate $SI\kappa$, where κ denotes the individual contact rate; a risky contact can result in a new infection with probability γ, $\bar{\gamma}$ denoting the mean of the transmission probabilities of all infected hosts at a given time; a host is removed from the infected compartment in the events of death (occuring at rate δ) or diagnosis (occurring at rate ρ); diagnosis is followed by immediate treatment, recovery, and lifelong immunity for the patient; healthy hosts leave the S and R compartments at a constant rate μ. (*C*) An example time-course of the trait value within a host—the value changes instantaneously with strain mutation; in the "neutral" case (black), the trait can change upward or downward; in the "select" case (magenta), only positive changes are possible (mutants resulting in a lower trait value can not substitute the current strain). (*D–F*) The per locus mutation rate (ν), the per risky contact transmission probability (γ) and the expected infectious time ($1/(\delta + \rho)$) are defined as constants in the "neutral" case (black) or as functions of the trait value in the "select" case (magenta) (supplementary table S2, Supplementary Material online).

are subject to change with a substitution of the dominant strain within the host (magenta curves on fig. 2*D–F*). For each class of events (within- and between-host), we define two modes:

- neutral: events occur at rates defined as global constants mimicking neutrality (i.e., lack of selection) with respect to z (black lines on fig. 2*C–E*). For within-host events, it is assumed that a mutation of the pathogen is followed by instantaneous substitution of the mutant for the current dominant strain, regardless of the induced change in z (black line on fig. 2*C*);
- select: within hosts, it is assumed that a mutation of the pathogen is followed by instantaneous substitution only if it results in a higher z (magenta line on fig. 2*C*). Borrowing the approach from (Fraser et al. 2007), the rates of transmission and within-host mutation are defined as increasing Hill functions of 10^z, whereas the infectious time period is defined as a decreasing Hill function of 10^z, thus mimicking increasing per capita

transmission- and pathogen-induced mortality for higher z (magenta lines on fig. 2*D–F*).

By combining different modes of dynamics at the within- and between-host levels the model can reproduce some popular hypotheses of pathogen evolution. For example, the combination of select within-host mode with select between-host mode simulates selection for optimal transmission potential (Fraser et al. 2007; Stearns and Koella 2007). This allows to evaluate the combined effect of selection and within-host trait evolution on various estimators of heritability.

Results

In this section, we use empirical data and simulations of the toy-model to show that most of the heritability estimators borrowed from classical quantitative genetics are prone to significant bias, because they neglect or inaccurately model the change in resemblance between transmission partners

caused by within-host evolution of the pathogen. Based on the toy-model simulations, we designate the intraclass correlation in the closest phylogenetic pairs (CPPs) and the phylogenetic heritability, $H^2_{OU}(\bar{t})$, measured by the phylogenetic Ornstein–Uhlenbeck mixed model (POUMM) (Mitov and Stadler 2016; Blanquart et al. 2017) as the most reliable estimators of pathogen trait heritability. Based on applying these estimators to a large HIV cohort, we establish a lower bound for the lg(spVL)-heritability.

Through the rest of the article, we use the symbol d_{ij} to denote the phylogenetic distance between two tips, i and j, on a transmission tree (fig. 1). d_{ij} summarizes the total evolutionary distance between two infected hosts at the moment of measuring the trait value and is measured in substitutions per site for real trees and arbitrary time units for simulated trees. We begin our report with a result from HIV data demonstrating the relevance of within-host evolution for estimating heritability.

The lg(spVL) Correlation in HIV Phylogenetic Pairs Decreases with d_{ij}

We used one-way analysis of variance (ANOVA, r_A) and Spearman correlation (r_{Sp}) to estimate the correlation in phylogenetic pairs (PP) extracted from a recently published transmission tree of 8,483 HIV patients (Hodcroft et al. 2014). As defined in Shirreff et al. (2013), phylogenetic pairs represent pairs of tips in the transmission tree that are mutually nearest to each other by phylogenetic distance (d_{ij}) (fig. 1). We ordered the PPs by d_{ij} and split them into ten strata of equal size (deciles), evaluating the correlation between pair trait values (r_A and r_{Sp}) in each stratum. The point estimates and the 95% confidence intervals (CI) are shown with black and magenta points and error bars on figure 3. Dashed horizontal bars denote the 95% CI for r_A evaluated on all phylogenetic pairs. Despite some irregularities, there is a well pronounced pattern of decay in the correlation—strata to the left (small d_{ij}) tend to have higher r_A values than strata to the right (big d_{ij}). The values of r_A closely matched the values from other correlation estimators, such as DR (b) and the Pearson product mean correlation (r) (results not shown). We performed ordinary least squares regressions (OLS) of the values r_{A,D_k} and r_{Sp,D_k} on the mean phylogenetic distance, $\bar{d}_{ij,k}$, in each stratum, $k = 1, \ldots, 10$. The slopes of both regressions were significantly negative ($P<0.05$) and are shown as black and magenta lines on figure 3. Similar slopes were obtained when using other stratifications of the data (supplementary fig. S1, Supplementary Material online).

The above result shows that the value of a heritability estimator based on the correlation within phylogenetic pairs (including DR couples) depends strongly on d_{ij}. Another issue of all estimators of H^2 using the correlation in phylogenetic or DR pairs is that the underlying statistical methods require independence between the pairs—the trait values in one pair should not influence or be correlated with the trait values in any other pair. This assumption is not valid in general, due to the phylogenetic relationship between all patients. One

way to mitigate the effects of phylogenetic relationship between pairs is to limit the analysis to the closest pairs (i.e., pairs, for which d_{ij} does not exceed some user specified threshold). This approach has the drawback of omitting much of the data from the analysis. As an alternative taking advantage of the entire tree, it is possible to correct for the phylogenetic relationship by using a phylogenetic comparative method (PCM). PCMs attempt to solve both of the above problems, because they 1) incorporate the branch lengths in the transmission tree to model the variance–covariance structure of the data and 2) correct for the phylogenetic correlation when estimating evolutionary parameters or the phylogenetic heritability of the trait (Felsenstein 1985; Housworth et al. 2004; Alizon et al. 2010). These advantages of the PCMs come at the price of assuming a specific stochastic process as a model of the trait evolution along the tree. In the next subsection, we show that assuming an inappropriate process for the trait evolution can cause a significant bias in the estimate of phylogenetic heritability.

A Brownian Motion Process Cannot Reproduce the Decay of Correlation in the UK Data

We implemented a maximum likelihood and a Bayesian fit of the PMM (Lynch 1991; Housworth et al. 2004) and its extension to an Ornstein–Uhlenbeck model of evolution (POUMM) (Hansen 1997; Mitov and Stadler 2016; Blanquart et al. 2017). The PMM and the POUMM assume an additive model of the trait values, $z(t) = g(t) + e$, in which $z(t)$ represents the trait value at time t for a given lineage of the tree, $g(t)$ represents a heritable (genotypic) value at time t for this lineage and e represents a nonheritable contribution summarizing the effects of the host and his/her environment on the trait and the measurement error. The only difference between the two models is their assumption about the evolution of $g(t)$ along the branches of the tree—the PMM assumes a Brownian motion process; the POUMM assumes an Ornstein–Uhlenbeck process (Uhlenbeck and Ornstein 1930; Lande 1976; Hansen 1997).

Using the maximum likelihood estimates of the model parameters (supplementary table S1, Supplementary Material online), we simulated random trait trajectories on the UK tree, running 100 replications for each model. For each replication, we estimated the correlation, r_A, in PPs using the simulated values instead of the real values. The resulting correlation estimates are shown on figure 3 as brown and green points and error bars for the PMM and POUMM simulations, respectively. We notice that there is a significant difference between the correlation estimates of the two models. In particular, in the leftmost decile the POUMM estimate is significantly higher than the PMM estimate (the POUMM 95% CI excludes the PMM estimate).

In order to understand the above difference between PMM and POUMM, we derive approximate analytical expressions of the correlation as a function of d_{ij} under the two models. Assume for simplicity that two tips i and j are situated at equal distance, t, from the root. According to Brownian motion (BM), the correlation is a function of t and the
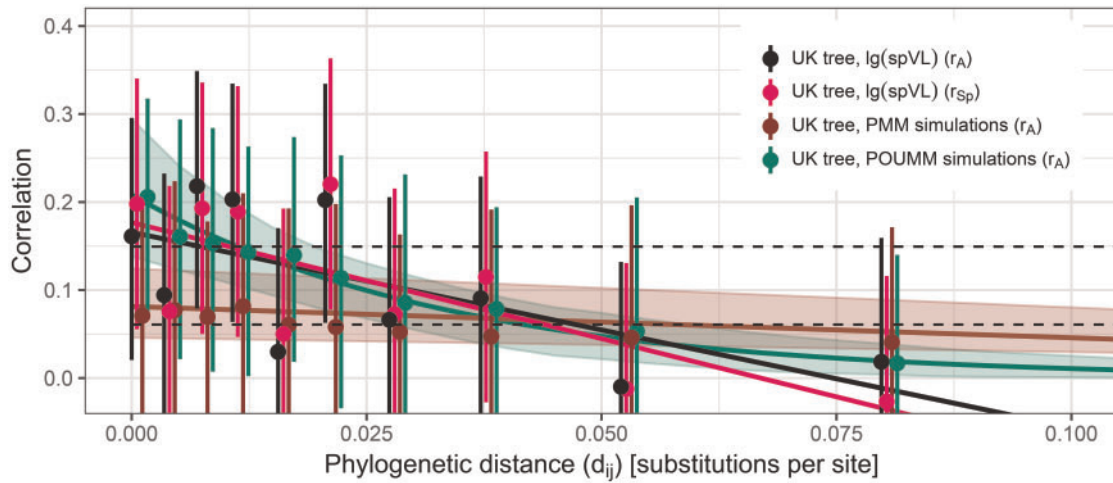
Fig. 3. Correlation between lg(spVL)-values in HIV phylogenetic pairs. A sample of 1917 PPs with lg(spVL)-measurements from HIV patients shows a decrease in the correlation (ICC) between pair trait values as a function of the pair phylogenetic distance d_{ij}. The point estimates and 95% CIs in ten strata of equal size (deciles) are depicted as points and error bars positioned at the mean d_{ij} for each stratum, $\bar{d}_{ij}$. Black and magenta points with error-bars denote the estimated r_A and r_{Sp} in the real data. Dashed horizontal bars denote the 95% CI for r_A evaluated on all phylogenetic pairs. A black and a magenta inclined line denote the least squares linear regression of r_A and r_{Sp} on $\bar{d}_{ij}$. Brown and green points with error bars denote the estimated values of r_A obtained after replacing the real trait values on the tree by values simulated under the maximum likelihood fit of the PMM and the POUMM methods, respectively (mean and 95% CI estimated from 100 replications). A brown and a green line show the expected correlation between pairs of tips at distance d_{ij}, as modeled under the ML-fit of the PMM and the POUMM (eqs. 2 and 3). A light-brown and a light-green region depict the 95% high posterior density (HPD) intervals inferred from Bayesian fit of the two models (Materials and Methods).

distance t_{ij} from the root to the pair's most recent common ancestor (mrca):

$$r_{BM,ij} = \frac{\text{Cov}_{BM}(t_{ij}; \sigma^2)}{\text{Var}_{BM}(t; \sigma^2) + \sigma_e^2}$$
$$= \frac{\sigma^2 t_{ij}}{\sigma^2 t + \sigma_e^2}, \tag{2}$$

where σ^2 denotes the unit time variance of the BM process and σ_e^2 denotes the variance of the environmental (nonheritable) component, e (Housworth et al. 2004, Materials and Methods). According to Ornstein–Uhlenbeck (OU), the correlation is a function of t, t_{ij}, as well as the phylogenetic distance between the tips, d_{ij}:

$$r_{OU,(ij)} = \frac{\text{Cov}_{OU}(t_{ij}, d_{ij}; \alpha, \sigma^2)}{\text{Var}_{OU}(t; \alpha, \sigma^2) + \sigma_e^2}$$
$$= \frac{\frac{\sigma^2}{2\alpha} \exp(-\alpha d_{ij})(1 - \exp(-2\alpha t_{ij}))}{\frac{\sigma^2}{2\alpha}(1 - \exp(-2\alpha t)) + \sigma_e^2}, \tag{3}$$

where the additional parameter α denotes the selection strength of the OU process (Hansen 1997). By plugging-in the ML estimates for the model parameters (supplementary table S1, Supplementary Material online), substituting t with the mean root-tip distance in the tree ($\bar{t} = 0.14$), and approximating t_{ij} with its linear regression on d_{ij} in the UK tree ($\hat{t}_{ij} = 0.15 - 0.63 d_{ij}$), we obtain:

$$r_{BM,ij} \approx 0.08 - 0.36 d_{ij}. \tag{4}$$

$$r_{OU,(ij)} \approx 0.21 \exp(-28.78 d_{ij}) \tag{5}$$
$$\times \left(1 - \underbrace{\exp(-8.35 + 36.47 d_{ij})}_{\approx 0}\right)$$
$$\approx 0.21 \exp(-28.78 d_{ij}).$$

The last approximation in equation (5) follows from the fact that the term $\exp(-8.35 + 36.47 d_{ij})$ is nearly 0 for the range of phylogenetic distances ($d_{ij} \in [0, 0.14]$) in the UK tree (see supplementary information, Supplementary Material online, for further details on the above approximations).

Equations (4) and (5) represent a linear and an exponential model of the correlation as a function of d_{ij}. The values of these equations at $d_{ij}=0$ are equal to the phylogenetic heritabilities estimated at the mean root-tip distance $\bar{t}$ under PMM and POUMM (details on that later). The slope of the linear model (eq. 4) equals -0.36 (95% HPD $[-0.58, -0.21]$). The rate of the exponential decay (eq. 5) equals the POUMM parameter $\alpha=28.78$ (95% HPD [16.64, 46.93]) and the half-life of decay equals $\ln(2)/\alpha = 0.02$ substitutions per site (95% HPD [0.01, 0.04]).

Plotting the values of equations (4) and (5) and their 95% HPD intervals on figure 3 reveals visually that the POUMM fits better to the data than the PMM. Statistically, this is confirmed by a lower Akaike Information Criterion (AICc) for the POUMM fit and a strictly positive HPD interval for the OU parameter α (supplementary table S1 and fig. S8, Supplementary Material online). The slope of the linear model derived from the PMM fit (eq. 4, brown line on fig. 3) is nearly flat compared with the slopes of the two OLS fits (black and magenta lines on fig. 3). To explain this,

we notice that in PMM, the covariance in phylogenetic pairs and the variance at the population level are modeled as linear functions of the root-mrca distance (t_{ij}) and the root-tip distance (t) (numerator and denominator in eq. 2). Importantly, both of these linear functions are bound to the same slope parameter, σ^2. As it turns out, in the UK data, the covariance and the variance increase at different rates with respect to t_{ij} and t (see supplementary fig. S2 and supplementary information, Supplementary Material online). We conclude that the PMM is not an appropriate model for the correlation in phylogenetic pairs, being unable to model the above difference in the rates.

In the limit $d_{ij} \rightarrow 0$, a phylogenetic pair should be equivalent to a DR couple at the moment of transmission, that is, before the genotypes in the two hosts have diverged due to within-host evolution. Thus, it appears reasonable to use an estimate of the correlation at $d_{ij}=0$ as a proxy for the broad-sense heritability, H^2, in the entire population. This idea has been applied in previous studies of HIV (Hecht et al. 2010; Hollingsworth et al. 2010; Bachmann et al. 2017; Blanquart et al. 2017) as well as malaria (Anderson et al. 2010). One potential obstacle to this approach is the possibility of introducing a sampling bias by filtering of the data. For example, if the study is on a trait, which evolves toward higher values during the course of infection, patients with lower trait values would tend to be more frequent among the CPPs than in the entire population. Thus, there is no guarantee that the trait distribution and, therefore, the heritability measured in the CPPs equals the heritability in the entire population. This problem of sampling bias affects both, resemblance-based as well as the currently used phylogenetic comparative methods. This suggests that the approach of imposing a threshold on d_{ij} or estimating the correlation (r_A, r_{Sp} or another correlation measure) at $d_{ij}=0$ needs further validation. In the next subsection, we use simulations of the toy model to show that sampling bias, although present, is comparatively small with respect to the negative bias due to measurement delay.

ANOVA-CPP and POUMM Are the Least Biased Heritability Estimators in Toy-Model Simulations
Here, we use simulations of the toy-model to compare a number of heritability estimators against the known true value of H^2 (measured directly by the coefficient of determination R^2_{adj}). We use the symbol T_{10k} to denote the transmission tree of the first 10,000 diagnosed individuals in a simulation. Below we list the different heritability estimators grouping them by the type of their input:

- *Grouping of the trait values by identical pathogen genotype.* We evaluated the coefficient of determination adjusted for finite sample size, R^2_{adj}, and the intraclass correlation (ICC) estimated using one-way ANOVA, $r_A[\text{id}]$. The main difference between these two estimators is the ANOVA assumption that the group-means (genotypic values) are sampled from a distribution of potentially many more genotypes than the ones found in the data. In contrast, R^2_{adj} assumes that all genotypes in the population are present in the sample. Since the latter

assumption is true for the simulated epidemics, R^2_{adj} represents the reference (true) value of H^2 to which all other estimates are compared.
- *Known DR couples.* We evaluated the regression slope of recipient on donor values in three ways: 1) b—based on the trait values at the moment of diagnosing the infection; 2) b_0—based on the trait values right after the transmission events; and 3) $b_{d_{ij}'}$—based on the subsample of diagnosed couples having d_{ij} not exceeding a threshold d_{ij}'. Based on a trade-off between precision and bias, we specified $d_{ij}' = D_1$, D_1 denoting the first decile in the empirical distribution of d_{ij} (see supplementary information, Supplementary Material online).
- *Phylogenetic pairs (PPs) in T_{10k}.* We evaluated ICC using ANOVA in three ways: 1) r_A—based on all PPs; 2) r_{A,D_1}—based on CPPs defined as PPs in T_{10k} having d_{ij} not exceeding the first decile, D_1; and 3) $r_{A,0,\text{lin}}$—the estimated intercept from a linear regression of the values r_{A,D_k} on the mean values $d_{ij,k}$ in each decile, $k = 1, \ldots, 10$; For the latter two estimators, which attempt to estimate r_A at $d_{ij}=0$, we use the acronym ANOVA-CPP. As an alternative to ANOVA, which is more robust to outliers (e.g., extreme values at the tails of the trait distribution), we evaluated the Spearman correlation in the first decile, hereby denoted as r_{Sp,D_1}.
- *Transmission tree T_{10k}.* We evaluated the phylogenetic heritability based on the ML fit of the PMM and POUMM models. Specifically, we compared the classical formula evaluated at the mean root-tip distance $\bar{t}$ in the tree (eqs. 10 and 12) (Housworth et al. 2004; Leventhal and Bonhoeffer 2016) and the empirical formula based on the sample trait variance, $s^2(z)$ (eqs. 11 and 13) (described in Materials and Methods). For the PMM, we denote these estimators by $H^2_{BM}(\bar{t})$ and H^2_{BMe}; for the POUMM, we use the symbols $H^2_{OU}(\bar{t})$ and H^2_{OUe}:

Table 1 summarizes the mathematical definition and the assumptions of the above estimators. A more detailed description of the PMM and the POUMM methods is provided in Materials and Methods. The referenced textbooks on quantitative genetics (Lynch and Walsh 1998) are excellent references for the other methods.

By combining "neutral" and "select" dynamics for the strain mutation and substitution rates at the within-host level, and the virus-induced per capita death rate and per contact transmission probability at the between-host level, we defined the following scenarios of the toy-model:

- Within: neutral/Between: neutral;
- Within: select/Between: neutral;
- Within: neutral/Between: select;
- Within: select/Between: select;

For each of these scenarios and mean contact interval $1/\kappa \in \{2, 4, 6, 8, 10, 12\}$ (arbitrary time units), we executed ten simulations resulting in a total of $4 \times 6 \times 10 = 240$ simulations. Of the 240 simulations, 175 resulted in epidemic outbreaks of at least 10,000 diagnosed hosts. For each

Table 1. Tested Estimators of the Broad-Sense Heritability of Pathogen Traits.

Input Data	Method (Abbreviation)	Assumptions	Estimator
Grouping by identical infecting strain	Adjusted coefficient of determination	The sample of data contains all genotypes present in the population	$R^2_{adj} = 1 - \frac{N-1}{N-K}\frac{s^2(z-\hat{G})}{s^2(z)}$ (6)
	One-way analysis of variance (ANOVA)	Independently sampled genotypes i.i.n.d. trait-values within each group	$r_A[id] = \frac{(M_{Sb}-M_{Se})/n}{(M_{Sb}-M_{Se})/n+M_{Se}}$ (7)
		Equal within-group variances (homoscedasticity)	
Known donor–recipient couples	Donor–recipient regression (DR)	Independently sampled donor–recipient couples	
		Equal residual variance across the range of donor-values (homoscedasticity)	$b = \frac{s(z_{don},z_{rcp})}{s^2(z_{don})}$, (8)
		Equal donor and population variances	variants: $b, b_0, b_{d_{ij}'}$
Phylogenetic pairs (PPs)	ANOVA on all/closest PPs (ANOVA-PP, ANOVA-CPP)	ANOVA assumptions (see above)	Defined as in equation (7), but calculated on PPs variants: $r_A, r_{A,d_{ij}'}$
	Spearman correlation on all/closest PPs	PPs are independent from one another	Pearson (product mean) correlation, calculated on the ranks of the trait-values. variants: $r_{Sp}, r_{Sp,d_{ij}'}$
	Linear regression of r_A on d_{ij} upon a stratification	r_A depends linearly on d_{ij}	The intercept, $r_{A,0,lin}$, from the OLS fit of the model
		Equal residual variance across the range of d_{ij}	$r_A(d_{ij}) = r_{A,0,lin} + \omega_1 d_{ij}$. (9)
Transmission tree	Phylogenetic mixed model (PMM)	Branching BM evolution	$H^2_{BM}(\bar{t}) = \bar{t}\sigma^2/(\bar{t}\sigma^2 + \sigma_e^2)$ (10)
		i.i.n.d. distributed environmental deviation, $e \sim N(0,\sigma_e^2)$	$H^2_{BMe} = 1 - \sigma_e^2/s^2(z)$ (11)
	Phylogenetic Ornstein–Uhlenbeck mixed model (POUMM)	Branching OU evolution	$H^2_{OU}(\bar{t}) = \frac{\sigma^2(1-\exp(-2\alpha\bar{t}))}{\sigma^2(1-\exp(-2\alpha\bar{t}))+2\alpha\sigma_e^2}$ (12)
		i.i.n.d. environmental deviation, $e \sim N(0,\sigma_e^2)$	$H^2_{OUe} = 1 - \sigma_e^2/s^2(z)$ (13)

NOTE.—Notation: $s^2(\cdot)$, sample variance; $s(\cdot,\cdot)$, sample covariance; N, number of patients; K, number of distinct groups of patients, that is, genotypes or phylogenetic pairs; z, measured values; $\hat{G}$, estimated genotypic values: mean values from patients carrying a given genotype; z_{don}, donor values; z_{rcp}, recipient values; M_{Se}, within-group mean square: $M_{Se} = \frac{\sum(z_i-\bar{z}_i)^2}{N-K}$, where z_i is an individual's value and $\bar{z}_i$ is the mean value of the group to which the individual belongs; M_{Sb}, among-group mean square: $M_{Sb} = \frac{\sum(\bar{z}_i-\bar{z})^2}{K-1}$, where $\bar{z}_i$ is defined as above and $\bar{z}$ is the population mean value; n, weighted mean number patients in a group, that is, $n=2$ for phylogenetic pairs and $n = \left(N - \frac{\sum n_i^2}{N}\right)/(K-1)$ for groups of variable size; α, σ, σ_e: PMM/POUMM parameters (described in Materials and Methods).
i.i.n.d., independent and identically normally distributed; d_{ij}, phylogenetic distance between donor–recipient pairs or phylogenetic pairs; d_{ij}', threshold on d_{ij} (see text).

outbreak, we analyzed the populations of the first up to 10,000 diagnosed hosts.

Rarer transmission events (bigger $1/\kappa$) result in longer transmission trees and, therefore, longer average phylogenetic distance between tips, d_{ij} (supplementary fig. S3, Supplementary Material online). This enabled demonstrating the effect of accumulating within-host evolution on the different heritability estimators (fig. 4).

Figure 4 shows that the estimators b_{D_1}, b, r_{A,D_1}, and r_A are negatively biased in general for all toy-model scenarios. This bias tends to increase with the mean contact interval, $1/\kappa$ (respectively, d_{ij}), because random within-host mutation tends to decrease the genetic overlap between DRs and phylogenetic pairs (supplementary fig. S4, Supplementary Material online). The negative bias was far less pronounced when imposing a threshold on d_{ij} but this came at the cost of precision (less biased but longer box-whisker plots for b_{D_1} and r_{A,D_1} compared with b and r_A) (fig. 4). Several additional

sources of bias were revealed when considering the practically unavailable estimators b_0 and $r_A[id]$. The estimator $r_A[id]$ was positively biased due to the small number of simulated genotypes (only six)—this was validated through additional simulations showing that $r_A[id]$ converges to the true value for a slightly bigger number of genotypes (e.g., $K \geq 24$ genotypes, see supplementary information, Supplementary Material online). The estimator b_0 was behaving accurately in the neutral/neutral scenario (excluding very short contact intervals) but tended to have a bias in both directions in all scenarios involving selection. The main reason for these biases was the phenomenon of "sampling bias" consisting in a difference between the distributions of measured values in the DR couples and the population of interest. Although its magnitude was comparetively small in the simulations, we presume that sampling bias could play an important role in real biological applications. We already gave an example of this bias in the previous subsection. Another manifestation of sampling bias
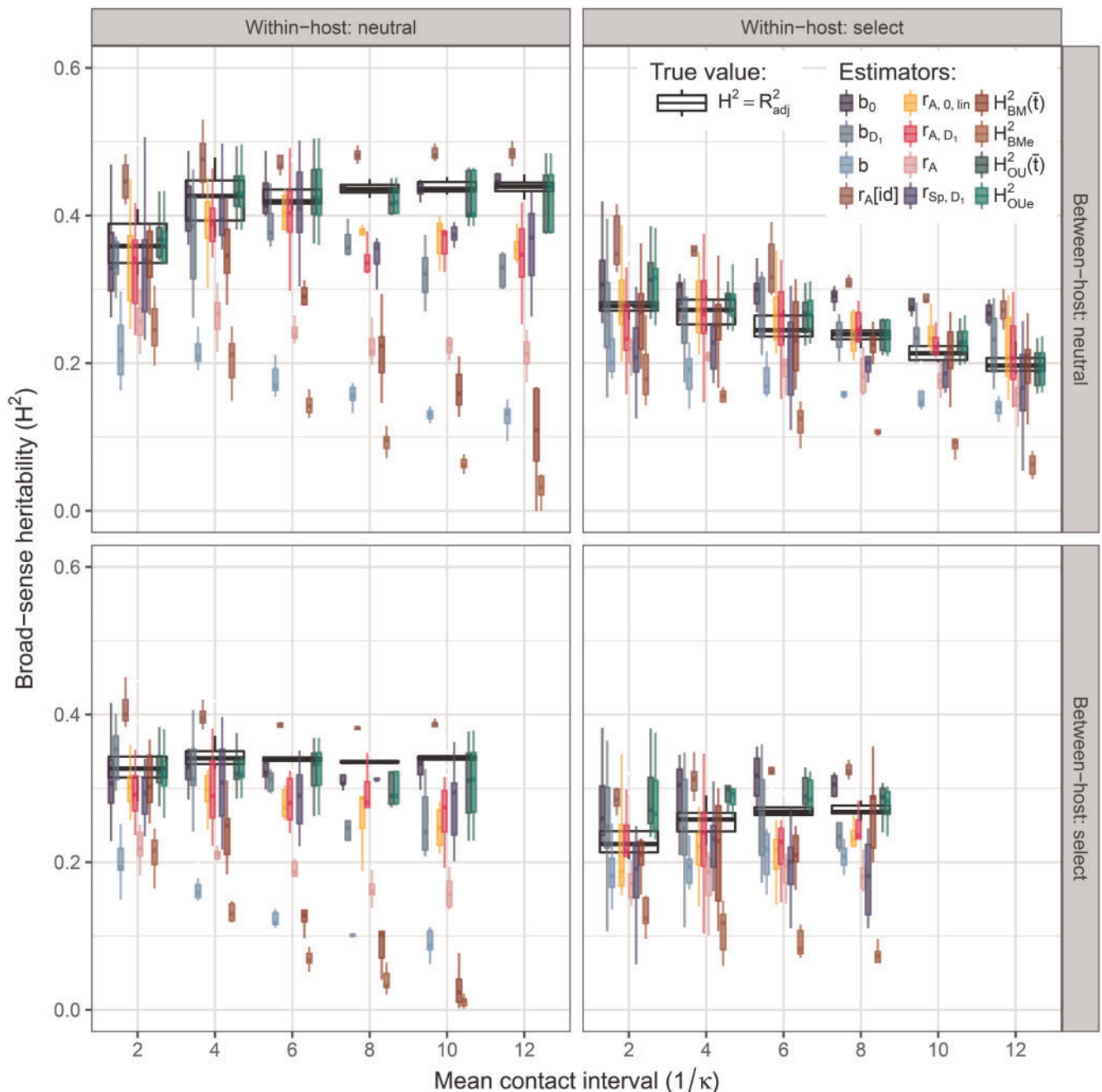
Fig. 4. Heritability estimates in toy-model simulations. (A–D) H^2-estimates in simulations of "neutral" and "select" within-/between-host dynamics. Each group of box-whiskers summarizes the simulations for a fixed scenario and contact interval, $1/\kappa$; white boxes (background) denote true heritability, colored boxes denote estimates (foreground). Statistical significance is evaluated through t-tests summarized in table 2.

is the fact that b_0 does not fully eliminate the effect of within host-evolution (and selection) in the donors. This is why, in cases of selection, the phenotypic variance in the donors tends to be smaller than the variance in the recipients as well as the variance in the population (supplementary fig. S5, Supplementary Material online). Additional details on these potential sources of bias are provided in supplementary information, Supplementary Material online.

Further, the simulations showed that a worsening fit of the BM model on longer transmission trees was causing an inflated estimate of the environmental deviation, σ_e, in the

PMM fits and, therefore, a negative bias in $H^2_{BM}(\bar{t})$ and H^2_{BMe} (compare estimates for small and big values of $1/\kappa$ on fig. 4 and supplementary fig. S6C, Supplementary Material online). In contrast with the PMM, the POUMM estimates, $H^2_{OU}(\bar{t})$ and H^2_{OUe} were far more accurate and the value of σ_e in the POUMM ML fit was nearly matching the true nonheritable deviation in most simulations (fig. 4 and supplementary fig. S6C, Supplementary Material online). The better ML fit of the POUMM was confirmed by stronger statistical support, namely by lower AICc values in all toy-model simulations (supplementary fig. S6D, Supplementary Material online).

The fact that the POUMM outperformed the PMM in all scenarios contradicted with the initial belief that the PMM should be the better suited model for a neutrally evolving trait represented by the neutral/neutral scenario, whereas the POUMM should fit better to scenarios involving selection. It was also counterintuitive that the inferred parameter α from the POUMM model was significantly positive in all simulations including the neutral/neutral scenario (supplementary fig. S6B, Supplementary Material online). To better understand this phenomenon, we performed the PP stratification analysis on the toy-model data (supplementary fig. S7, Supplementary Material online). This revealed a pattern of correlation that decays exponentially with d_{ij}. The shape of exponential decay was mostly pronounced for longer contact intervals, $1/\kappa$, particularly in the neutral/neutral scenario (first column on supplementary fig. S7, Supplementary Material online). In supplementary information, Supplementary Material online, we show that an exponentially decaying phenotypic correlation is consistent with a neutrally mutating genotype under a Jukes–Cantor substitution model (Yang 2006). The decay of the correlation was still present in scenarios involving within- and/or between-host selection but the observed pattern was rather irregular and deviating from an exponential function of d_{ij} (supplementary fig. S7, Supplementary Material online). In most cases, the ML fit of the PMM method was a bad fit to the decay of correlation (brown dots and error-bars on supplementary fig. S7, Supplementary Material online); for longer contact intervals, there was a tendency toward constant values of the correlation under PMM far below the true value (brown dots and error bars on supplementary fig. S7, Supplementary Material online). This explains the overall better accuracy of the POUMM versus the PMM method.

Table 2 shows the average bias of each tested estimator for each of the four scenarios. We conclude that, apart from the practically inaccessible estimators based on grouping by identical genotype (R^2_{adj} and $r_A[id]$), the most accurate estimators of H^2 in the toy-model simulations are $H^2_{OU}(\bar{t})$ and H^2_{OUe} followed by estimators of the correlation in PPs minimizing the phylogenetic distance d_{ij}, that is (r_{A,D_1}, $r_{A,0,lin}$, r_{Sp,D_1}). In the next subsection, we report the results from these estimators in the UK HIV data.

Heritability of lg(spVL) in the UK HIV Cohort

We evaluated the correlation in the CPPs (ANOVA and Spearman correlation) in data from the UK HIV cohort comprising lg(spVL) measurements and a tree of viral (*pol*) sequences from 8,483 patients inferred previously in (Hodcroft et al. 2014). In addition, we performed a Bayesian fit of the POUMM and the PMM methods to the same data. The goal was to test our conclusions on a real data set and to compare the H^2-estimates from CPPs and POUMM to previous PMM/ReML-estimates on exactly the same data (Hodcroft et al. 2014).

In applying ANOVA-CPP, the first step has been to define the threshold phylogenetic distance for defining CPPs. To that end, we explored different stratifications of the PPs as shown on supplementary figure S1B, Supplementary Material online, and a scatter plot of the phylogenetic distances against the

Table 2. Mean Difference $\widehat{H^2} - R^2_{adj}$ from the Toy-Model Simulations Grouped by Scenario.

Within: Between:	Neutral Neutral	Neutral Select	Select Neutral	Select Select
N	50	41	47	37
b_0	−0.01*	−0.02**	0.05**	0.04**
b_{D_1}	−0.07**	−0.04**	0	−0.01
b	−0.25**	−0.2**	−0.07**	−0.06**
$r_A[id]$	0.05**	0.05**	0.08**	0.06**
$r_{A,0,lin}$	−0.05**	−0.06**	0.01	−0.04**
r_{A,D_1}	−0.05**	−0.06**	0	−0.03*
r_A	−0.18**	−0.15**	−0.06**	−0.08**
r_{Sp,D_1}	−0.05**	−0.05**	−0.05**	−0.07**
$H^2_{BM}(\bar{t})$	−0.17**	−0.17**	−0.01	−0.04*
H^2_{BMe}	−0.28**	−0.24**	−0.12**	−0.16**
$H^2_{OU}(\bar{t})$	−0.01	−0.02**	0.01*	0.03**
H^2_{OUe}	−0.01	−0.02**	0.01*	0.03**

NOTE.—Statistical significance is estimated by Student's t-tests, P values denoted by an asterisk as follows: * $P<0.01$; ** $P<0.001$. Gray background indicates estimates that are unavailable in practice.

absolute phenotypic differences, $|\Delta \lg(\text{spVL})|$ (fig. 5A). This revealed a small set of 116 PPs having $d_{ij} \leq 10^{-4}$ and narrowly coinciding with the first vigintile (also called 20-quantile or ventile) of d_{ij}. The phylogenetic distance in all remaining tip-pairs was more than an order of magnitude bigger, that is, $d_{ij} > 10^{-3}$. Given that the phylogenetic distance on the transmission tree is measured in substitutions per site and the length of the *pol*-region is in the order of 10^3 sites, we presume that the above set of 116 PPs corresponds to a set of 116 pairs of identical *pol* consensus sequences (no sequence data were available to check this). Based on this observation, we defined the above pairs as CPPs and the threshold was formally set to $d_{ij}' = 10^{-4}$. We validated that the CPPs were randomly distributed along the tree (fig. 5B). The random distribution of the CPPs along the transmission tree suggests that these phylogenetic pairs correspond to randomly occurring early detections of infection (trait values from each pair depicted as magenta segments on fig. 5B). To check that the filtering of the data, did not introduce a considerable sampling bias due to selection (see previous subsection), we also validated that there was no substantial difference in the trait distributions of all patients, the PPs and the CPPs (fig. 5C).

We compared the following estimators of H^2:

- ANOVA-CPPs (r_{A,D_1}, $r_{A,10^{-4}}$, r_{A,V_1}) and the original PP-method r_A;
- The intercept from the linear regression of r_A on d_{ij} upon a stratification of the PPs into deciles ($r_{A,0,lin}$, eq. 9);
- Spearman correlatoin in CPPs (r_{Sp,D_1}, $r_{Sp,10^{-4}}$, r_{Sp,V_1}) and in all PPs (r_{Sp}).
- The intercept from the linear regression of r_{Sp} on d_{ij} upon a stratification of the PPs into deciles ($r_{Sp,0,lin}$);
- POUMM ($H^2_{OU}(\bar{t})$, H^2_{OUe}), versus PMM ($H^2_{BM}(\bar{t})$, H^2_{BMe}) on the entire tree;

The results from these analyses are reported in table 3. ANOVA- and Spearman-correlation estimates, which
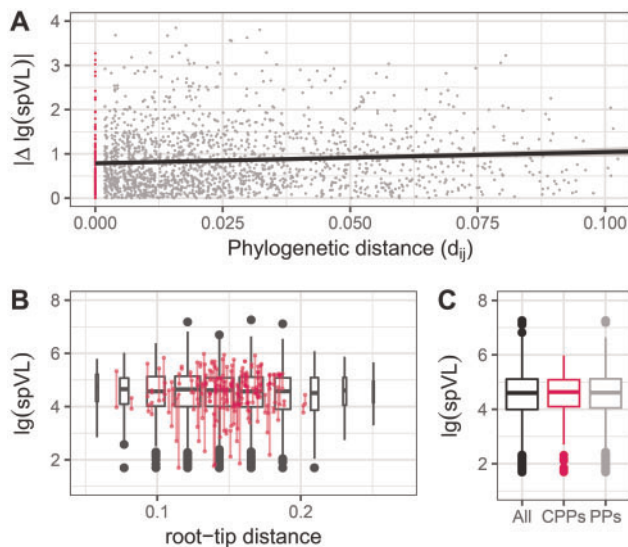
Fig. 5. Phylogenetic pairs in lg(spVL) data from the United Kingdom. (A) A scatter plot of the phylogenetic distances between pairs of tips against their absolute phenotypic differences: gray, PPs ($d_{ij} > 10^{-4}$); magenta, CPPs ($d_{ij} < 10^{-4}$). A black line shows the linear regression of $|\Delta \text{lg(spVL)}|$ on d_{ij} (the slope of the regression was statistically positive at the 0.01 level). (B) A box-plot representing the trait-distribution along the transmission tree. Each box-whisker represents the lg(spVL)-distribution of patients grouped by their distance from the root of the tree measured in substitutions per site. Wider boxes indicate groups bigger in size. Segments in magenta denote lg(spVL)-values in CPPs. (C) A box-plot of the lg(spVL)-distribution in all patients (black), PPs (gray), and CPPs (magenta).

minimized the phylogenetic distance by means of regression or filtering of the phylogenetic pairs had point-estimates of $r_{A,10^{-4}} = 0.17$ and $r_{Sp,10^{-4}} = 0.22$. The slightly higher estimate for the Spearman correlation could be explained by the presence of outliers in the data. Applying the POUMM to the entire tree reported a point estimate $H^2_{OU}(\bar{t}) = 0.21$ (8,483 patients, 95% CI [0.14, 0.29]).

Conversely, the heritability estimates from the original PP method (ANOVA or Spearman correlatoin on all PPs) and the PMM were significantly lower and falling below the 95% CIs from the POUMM (table 3). This confirms the observation from the toy-model simulations that these estimators are negatively biased, since they ignore or inaccurately model the changing correlation within pairs of tips. We validated the stronger statistical support for the POUMM with respect to the PMM, by its lower AICc value (supplementary table S1, Supplementary Material online) and by the posterior density for the POUMM parameter α (supplementary fig. S8, Supplementary Material online).

Finally, we compared our estimates of lg(spVL)-heritability to previous applications of the same methods on different data sets (fig. 6). In agreement with the toy-model simulations, estimates of H^2 using PMM or other BM-based phylogenetic methods (i.e., Blomberg's K and Pagel's λ) are notably lower than all other estimates, suggesting that these phylogenetic comparative methods underestimate H^2; resemblance-based estimates are down-biased by

measurement delays (e.g., compare early vs. late in the Netherlands on fig. 6).

In summary, POUMM and ANOVA-CPP yield agreeing estimates for H^2 in the UK data and these estimates agree with resemblance-based estimates in data sets with short measurement delay (different African countries and the Netherlands). Similar to the toy-model simulations, we notice a well-pronounced pattern of negative bias for the other estimators, PMM and ANOVA-PP, as well as for the previous resemblance-based studies on data with long measurement delay.

Discussion

Clarifying the Terminology and Notation

In this study, we explored how the differences between pathogens and mating species affect the various tools employed in estimating the heritability of pathogen traits. For mating species, the resemblance between relatives has been directly associated with the genetic determination of quantitative traits. The most prominent example is the parent–offspring regression slope used to estimate the narrow-sense heritability, h^2. For pathogens, one needs to disentangle the concepts of resemblance and genetic determination. First of all, the only reason to associate the parent–offspring regression slope with narrow-sense heritability is the presence of genetic segregation and recombination during sexual reproduction, favoring the inheritance of single-locus additive effects over multilocus epistatic effects (Lynch and Walsh 1998). Given that clonal pathogen transmission excludes segregation and recombination, the above association is invalid for pathogen traits. The correlation between transmission partners should rather be associated with the broad-sense heritability, H^2. This association, though, is compromised by a number of sources of bias, such as partial quasispecies transmission, within-host evolution, and many potential cofactors, such as shared habitats between donors and recipients, sampling bias, and convergent within-host evolution. All methods reviewed in this article can be regarded as methods that estimate the correlation between patients infected with identical pathogen strains. This is true also for the phylogenetic approaches, since, technically, the phylogenetic heritability is the expected correlation between pairs of tips in the limit $d_{ij} \to 0$ (see also Materials and Methods). Thus, all estimators can only be regarded as statistics summarizing the resemblance that is still observable in the presence of the above factors.

A Disagreement between Simulation Studies

Using simulations of the toy epidemiological model, we have shown that two methods based on phenotypic and sequence data from patients—estimating the correlation in CPPs and fitting the POUMM to the data—provide more accurate heritability estimates compared with previous approaches like DR and PMM. However, we should not neglect the arising discrepancy between our and previous simulation reports advocating either PMM (Hodcroft et al. 2014) or DR (Leventhal and Bonhoeffer 2016) as unbiased heritability

Table 3. Estimates of lg(spVL)-Heritability in HIV Data from the United Kingdom.

Method	N	$\hat{H}^2$	95% CI	95% HPD
Linear regression of r_A on $\bar{d}_{ij}$ in deciles (eq. 9) ($r_{A,0,lin}$)	10 points	0.17	[0.09, 0.24]	–
Linear regression of r_{Sp} on $\bar{d}_{ij}$ in deciles ($r_{Sp,0,lin}$)	10 points	0.18	[0.11, 0.25]	–
ANOVA-CPP (r_{A,V_1})	224	0.17	[−0.02, 0.31]	–
ANOVA-CPP ($r_{A,10^{-4}}$)	232	0.16	[0.01, 0.30]	–
ANOVA-CPP (r_{A,D_1})	384	0.16	[0.06, 0.25]	–
ANOVA-PP (r_A)[a]	3,834	0.11	[0.08, 0.14]	–
Spearman-CPP (r_{Sp,V_1})	224	0.23	[0.05, 0.42]	–
Spearman-CPP ($r_{Sp,10^{-4}}$)	232	0.22	[0.03, 0.4]	–
Spearman-CPP (r_{Sp,D_1})	384	0.2	[0.06, 0.34]	–
Spearman-PP (r_{Sp})[a]	3,834	0.11	[0.06, 0.15]	–
POUMM ($H^2_{OU}(\bar{t})$)	8,483	0.21	–	[0.14, 0.29]
POUMM (H^2_{OUe})	8,483	0.2	–	[0.13, 0.29]
PMM ($H^2_{BM}(\bar{t})$)[b]	8,483	0.08	–	[0.05, 0.12]
PMM (H^2_{BMe})[b]	8,483	0.06	–	[0.02, 0.1]
PMM, ReML (Hodcroft et al. 2014)[b]	8,483	0.06	[0.03, 0.09]	–

NOTE.—Also written are the results from a previous analysis on the same data set (Hodcroft et al. 2014). "–": the analysis was not done in the mentioned study. Gray background: estimates considered unreliable due to: [a]negative bias caused by measurement delays and [b]negative bias caused by BM violation. Uncertainty in the estimates is expressed in terms of 95% confidence intervals (CI), or, in the case of Bayesian inference, by 95% high posterior density intervals (HPDs).

estimators. Both of these studies have modeled within-host evolution, but failed to demonstrate the biases shown in this article. This could be explained by simulation artifacts. Hodcroft et al. (2014) perform simulations under a PMM model, so it is unlikely to reveal any bias in the PMM estimator; Leventhal and Bonhoeffer (2016) evaluated DR in consecutive Wright–Fisher generations using the donor values at the moment of transmission, thus, excluding potential measurement delay in the donors and accounting for a minute measurement delay in the recipients (one generation on the scale of hundreds of simulated generations). Compared with these simulations, the toy-model presented here has several important advantages: 1) it is biologically motivated by phenomena such as pathogen sequence mutation during infection, transmission of entire pathogens instead of proportions of trait values, and within-/between-host selection; 2) it allows to compare various resemblance-based and phylogenetic heritability estimates against the direct estimator, R^2_{adj}; 3) it is a fair test for all estimators of heritability, because it does not obey any of the estimators' assumptions, such as linearity of recipient—on donor values, normality of trait values, OU or BM evolution, independence between pathogen and host effects; and 4) it generates transmission trees that reflect the between-host dynamics, for example, clades with higher trait values exhibit denser branching in cases of between-host selection. As a criticism, we note that the toy-model does not allow strain coexistence within a host and, thus, is not able to model partial quasispecies transmission and, in particular, transmission bottlenecks (Keele et al. 2008) or preferential transmission of founder strains (Lythgoe and Fraser 2012). Although it may be exciting from a biological point of view, the inclusion of strain coexistence comes with a series of conceptual challenges, such as the definition of genotype and clonal identity or the formulation of the trait value as a function of a quasispecies—instead of a single strain genotype. These challenges should be addressed in future studies

implementing more advanced models of within-host dynamics and leveraging deep sequencing data. To conclude, the discrepancy between simulation studies highlights that no inference method suits all simulation setups *ergo* biological contexts. Thus, rather than proving universality of a particular method, simulations should be used primarily to study how particular biologically relevant features affect the methods on the table.

The Heritability of HIV Set-Point Viral Load Is at Least 20%

Applied to data from the United Kingdom, POUMM reported three times higher point estimates and nonoverlapping HPDs compared with a previous PMM/ReML-based estimate on the same data (0.06, 95% CI [0.02, 0.09]) (Hodcroft et al. 2014). Our PMM implementation confirmed this estimate. However, based on figure 3 and our simulations (fig. 4), the PMM estimates are underestimates of the true heritability. The estimate of 20% should still be considered a lower bound since it does not account for additional sources of potential negative bias, such as partial quasispecies transmission and measurement error. This result matches estimates from GWAS studies on the pathogen revealing that genetic polymorphisms in the virus explain ∼20% from spVL variance in other cohorts (reviewed in Bonhoeffer et al. 2015). Overall, our analyses yield an unprecedented agreement between estimates of DR resemblance and phylogenetic heritability in large European data sets and African cohorts, provided that measurements with large delays have been filtered out prior to resemblance evaluation (Hecht et al. 2010; Hollingsworth et al. 2010) (fig. 6A). Also noteworthy are the facts that our estimates for the UK data set support the results from Fraser et al. (2014) who conducted a meta-analysis of three data sets on known transmission partners (Hollingsworth et al. 2010; Lingappa et al. 2013; Yue et al. 2013) (433 pairs in total) reporting heritability values of
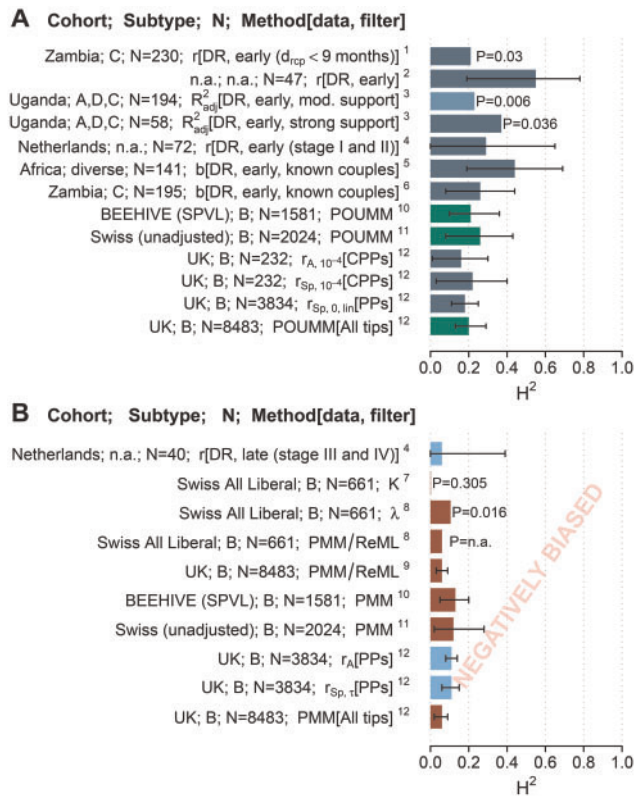
FIG. 6. A comparison between H^2H^2-estimates from the UK HIV-cohort and previous estimates on African, Swiss, and Dutch data. (A) Estimates with minimized measurement delay (dark cadet-blue) and POUMM estimates (green); (B) Down-biased estimates due to higher measurement delays (light-blue) or violated BM-assumption (brown). Confidence is depicted either as segments indicating estimated 95% CI or P values in cases of missing 95% CIs. References to the corresponding publications are written as numbers in superscript as follows: 1: Tang et al. (2004); 2: Hecht et al. (2010); 3: Hollingsworth et al. (2010); 4: van der Kuyl et al. (2010); 5: Lingappa et al. (2013); 6: Yue et al. (2013); 7: Alizon et al. (2010); 8: Shirreff et al. (2013); 9: Hodcroft et al. (2014); 10: Blanquart et al. (2017); 11: Bertels et al. (2018); 12: this work. For clarity, estimates from previous studies, which are not directly comparable (e.g., previous results from Swiss MSM/strict data sets; Alizon et al. 2010).

0.33, CI [0.20, 0.46], as well as the recent results from Blanquart et al. (2017) who conducted a POUMM and a PMM analysis on a whole-genome meta-data set (1,581 sequences from several European countries) reporting spVL heritability of 0.31, CI [0.15, 0.43]. In analogy with our ANOVA approach, Blanquart et al. (2017) measured the Pearson correlation in "cherries" partitioned by phylogenetic distance, showing a similar pattern of decreasing correlation with d_{ij}. Contrary to the UK data though, Blanquart et al. (2017) have shown nearly equal statistical support for PMM ($\alpha=0$, AIC = 3,343.2) and POUMM ($\alpha=7.6$, 95% bootstrap CI [1.2, 10.0], AIC = 3,344.5) for 1,581 subtype B *pol* sequences and spVL measurements (table 1 in Blanquart et al. 2017). This equal support fot the PMM and the POUMM models might indicate that none of the two models is a good fit to the data (i.e., flat likelihood surface), or that the likelihood surface for the POUMM is bimodal with modes at $\alpha=0$ and at $\alpha=7.6$. A

Bayesian POUMM fit with uninformative prior could be used to reveal such anomalies (see Materials and Methods and supplementary fig. S8, Supplementary Material online).

To sum up, all data sets support the hypothesis of HIV influencing spVL ($H^2>0.2$). The particular estimates provided here should be interpreted as lower bounds for H^2, because the partial quasispecies transmission, the noises in spVL measurements and the noise in transmission trees are included implicitly as environmental (nontransmittable) effects. The nonzero heritability motivates further HIV whole-genome sequencing (Metzner 2016) and genome-wide studies of the viral genetic association with viral load and virulence.

A Critical View on the POUMM

The OU process has found previous applications as a model for stabilizing selection in macroevolutionary studies (Lande 1976; Felsenstein 1988; Hansen 1997; Hansen and Bartoszek 2012) and references therein. As a contribution of this work, we have shown that the OU process is well adapted for the modeling of pathogen evolution along transmission trees in both, neutral as well as selection scenarios. The key advantage of the OU process to the BM process is the way in which the phylogenetic distance between a pair of tips enters in the expression for their correlation (eq. 3). This is a crucial advantage in modeling the loss of resemblance caused by within-host evolution of the pathogen (fig. 3 and supplementary fig. S7, Supplementary Material online). But there is a caveat coming along with this property of the OU-model—both, the rate at which a trait evolving under OU adapts toward θ and the rate of correlation decay for a pair of tips are governed by the same parameter: α. This is why a significantly positive estimate for α does not necessarily imply stabilizing selection. This was clearly shown in the neutral/neutral scenario of the toy-model simulations (supplementary fig. S6B, Supplementary Material online). A further extension of the POUMM using two separate parameters for the rate of attraction toward θ and for the rate of decorrelation would allow to disentangle the two forces.

Most of the above-mentioned studies and the accompanying software packages implementing phylogenetic OU models have assumed that the whole trait evolves according to an OU process, usually disregarding the presence of a biologically relevant nonheritable component e or treating it as a measurement error whose variance is a priori known (FitzJohn 2012). Having the OU process act on the genotypic values rather than whole trait values is a simplifying assumption facilitating mathematical processing (Mitov and Stadler 2016). However, our toy model simulations have shown robustness and statistical power of the POUMM in complicated scenarios combining trait-based selection at the within- and between-host levels.

A last criticism that can be addressed to the POUMM method is that it is unaware of between-host selection and demographic processes, which may result in a correlation between tree structure and trait values (e.g., higher branching density in clades with higher z). As noted by Leventhal and Bonhoeffer (2016), this is a general issue with phylogenetic comparative approaches assuming a global evolutionary

process acting on the whole phylogeny. An unexplored alternative would be to associate different instances of POUMM to different clades in the tree based on prior knowledge about heterogeneity between these clades.

Outlook

ANOVA-CPP and POUMM have great potential to become widely used tools in the study of pathogens. The accompanying R-package patherit provides a common interface for using the two methods on a transmission tree and phenotype data (Materials and Methods). ANOVA-CPP works on pairs of trait values from carriers of nearly identical strains and can be easily extended to groups of variable size (Lynch and Walsh 1998; Anderson et al. 2010). Thus, ANOVA-CPP is ideal for slowly evolving pathogens such as DNA-viruses, bacteria, and protozoa, where clusters of patients carrying identical-by-descent (IBD) strains are frequently found. For example, Anderson et al. (2010) identified 27 clusters of two to eight carriers of IBD strains in a small set of 185 malaria patients, that is, 41% of the patients participated in clusters. On the other hand, IBD-pairs are rare for rapidly evolving RNA-viruses, such as HIV and HCV. For instance, we identified only 116 CPPs in a large data set of 8,483 HIV-sequences, that is, <3% of the patients involved in IBD-pairs. However, the rapidly accumulating sequence diversity of RNA-viruses allows building large-scale phylogenies, which approximate transmission trees between patients. Thus, RNA-viruses should make the ideal scope for the POUMM. If the transmission tree is large enough, it is be possible to compare the estimates from the two methods and to analyze the profile of the correlation in phylogenetic pairs, as we did in the UK HIV data (fig. 3 and supplementary fig. S2, Supplementary Material online). We believe that, together, the two methods enable accurate and robust heritability estimation in a broad range of pathogens.

Materials and Methods

The subsections below provide details on the different heritability estimators (based on the categorization by input type, table 1) and the toy-model simulations.

Grouping by Identical Infecting Strain
Adjusted Coefficient of Determination
We calculated R^2_{adj} based on equation (6) (table 1).

One-Way Analysis of Variance
We calculated r_A based on equation (7) (table 1). A more detailed description of one-way ANOVA can be found in chapter 18 of Lynch and Walsh (1998).

Donor–Recipient Couples
To calculate the DR regression slope (b, b_0, b_{D_1}), we used equation (8) (table 1).

Phylogenetic Pairs
To calculate ICC in phylogenetic pairs (r_A, r_{A,D_1}, r_{A,V_1}, $r_{A,10^{-4}}$), we used one-way ANOVA (eq. 7, chapter 18 of Lynch and Walsh 1998). To calculate confidence intervals for the HIV

data, we used the R-package "boot" to perform 1,000-replicate bootstraps, upon which we called the package function boot.ci() with type="basic." These confidence intervals were fully contained in the standard ANOVA confidence intervals, based on the F-distribution (Lynch and Walsh 1998), which were slightly wider (not reported).

Phylogenetic Methods
Phylogenetic Mixed Model
The PMM assumes an additive model $z(t) = g(t) + e$, in which $z(t)$ represents the trait value at time t for a given lineage of the tree, $g(t)$ represents a heritable (genotypic) value at time t for this lineage and e represents the environmental (nonheritable) contribution. The genotypic value, $g(t)$, is assumed to evolve according to a branching Brownian motion process defined by the stochastic differential equation:

$$dg(t) = \sigma dW_t,$$
$$g(0) = g_0 \tag{14}$$

where g_0 is the initial genotypic value at the root, W_t is the standard Wiener process, and $\sigma > 0$ is the unit-time SD (Grimmett and Stirzaker 2001).

The environmental contribution e can change along the tree in any way as long as the values e at the tips are independent and identically normally distributed (i.i.n.d.) with mean 0 and variance σ_e^2. In the case of modeling an epidemic, e represents the total contribution from the host immune system, other host factors (e.g., age, sex), the host environment and measurement error; it obtains a value at the beginning of an infection, which can stay constant or change during the course of an infection, but is uncorrelated to the immune system and cofactors of other hosts.

Phylogenetic Ornstein–Uhlenbeck Mixed Model
The POUMM is an extension of the PMM replacing the BM assumption with an assumption of an Ornstein–Uhlenbeck (OU) process for the genotype evolution. The OU-process represents a continuous time random walk, which tends to move around a long-term mean value with greater attraction when the process is further away from that value (Uhlenbeck and Ornstein 1930; Hansen 1997). Technically, this is accomplished by adding an attraction term to equation (14):

$$dg(t) = \underbrace{\alpha[\theta - g(t)]dt}_{\text{Attraction to } \theta} + \underbrace{\sigma dW_t}_{\text{Brownian motion}}, \tag{15}$$

where θ denotes the long-term mean and $\alpha > 0$ is the attraction strength. Since in the limit $\alpha \to 0$ the attraction term vanishes and only the BM term remains, the OU-process represents a generalization of BM. As in the PMM, an independent white noise term $e \sim \mathcal{N}(0, \sigma_e^2)$ is added to $g(t)$ at the tips.

Phylogenetic Heritability
Introduced as a term with the PMM method (Housworth et al. 2004), the phylogenetic heritability quantifies how much

of the trait variance is attributable to g based on a fit of the assumed evolutionary model (in this case, BM or OU). For the BM and the OU processes, the genotypic variance is a function of the model parameters and the time-distance from the root of the ultrametric tree, t (Hansen 1997; Housworth et al. 2004):

$$\text{Var}_{\text{BM}}(t;\ \sigma) = \sigma^2 t \qquad (16)$$

$$\text{Var}_{\text{OU}}(t;\ \alpha, \sigma) = \frac{\sigma^2}{2\alpha}(1 - \exp(-2\alpha t)). \qquad (17)$$

Given the assumption that g and e are uncorrelated, the phenotypic variance is the sum of the genotypic variance and σ_e^2. Therefore, the phylogenetic heritability is also a function of t:

$$H_{\text{BM}}^2(t;\ \sigma, \sigma_e) = \frac{\text{Var}_{\text{BM}}(t;\ \sigma)}{\text{Var}_{\text{BM}}(t;\ \sigma) + \sigma_e^2} = \frac{\sigma^2 t}{\sigma^2 t + \sigma_e^2}, \qquad (18)$$

$$
\begin{aligned}
H_{\text{OU}}^2(t;\ \alpha, \sigma, \sigma_e) &= \frac{\text{Var}_{\text{OU}}(t;\ \alpha, \sigma)}{\text{Var}_{\text{OU}}(t;\ \alpha, \sigma) + \sigma_e^2} \\
&= \frac{\frac{\sigma^2}{2\alpha}(1 - \exp(-2\alpha t))}{\frac{\sigma^2}{2\alpha}(1 - \exp(-2\alpha t) + \sigma_e^2.}
\end{aligned}
\qquad (19)
$$

The above dependency of H_{OU}^2 and H_{BM}^2 on time is posing a problem in the case of a nonultrametric transmission tree, because the tips are at different time-distance from the root and do not share the same genotypic and phenotypic variance. We tested two possible work arounds: 1) evaluating the heritability at the mean root-tip distance, $\bar{t}$ (Leventhal and Bonhoeffer 2016); and 2) using an empirical definition of the phylogenetic heritability based on the empirical variance in the observed population:

$$H_e^2 = 1 - \frac{\sigma_e^2}{s^2(z)}. \qquad (20)$$

PMM and POUMM Log-Likelihood

The PMM and the POUMM log-likelihood represents the log-probability density of the observed data at the tips of the tree for given values of the model parameters, Θ. For PMM, $\Theta = <g_0, \sigma, \sigma_e>$; for POUMM $\Theta = <g_0, \alpha, \theta, \sigma, \sigma_e>$. Given that the two models are Gaussian, the log-likelihood is defined as the Gaussian log-probability density function:

$$
\begin{aligned}
\ell\ell(\Theta) = \ln f(\mathbf{z}|\Theta) = &-\frac{1}{2}\left(N\ln(2\pi) + \ln|\mathbf{V}_\Theta| + \right. \\
&\left. (\mathbf{z} - \boldsymbol{\mu}_\Theta)'\mathbf{V}_\Theta^{-1}(\mathbf{z} - \boldsymbol{\mu}_\Theta)\right),
\end{aligned}
\qquad (21)
$$

where $\mathbf{z}$ is the observed vector of trait values at the tips, $\boldsymbol{\mu}_\Theta$ is the mean vector at the tips ($\mu_i = g_0$ in the case of BM; $\mu_i = \exp(-\alpha t_i)g_0 + (1 - \exp(-\alpha t_i))\theta$ in the case of OU), and $\mathbf{V}_\Theta$ is the variance covariance matrix with off-diagonal elements given by the nominators and diagonal elements given by the denominators in equations (2) and (3), respectively.

PMM and POUMM Inference in the Toy-Model Simulations

The POUMM and PMM inference was done using maximum likelihood (ML) fit.

PMM and POUMM Inference on HIV Data

For HIV data, in addition to an ML-fit, we performed a Bayesian (MCMC) fit using an adaptive Metropolis algorithm with coerced acceptance rate (Vihola 2012) written in R (Scheidegger 2012).

The MCMC sampling was performed on the parameters g_0, α, θ, $H^2(\bar{t})$ and σ_e^2 (for likelihood and posterior density calculation, the parameter σ^2 was mapped back from $H^2(\bar{t})$ according to eqs. 18 and 19). The prior was specified as a joint distribution of independent variables: $(g_0,\ \alpha,\ \theta,\ H^2(\bar{t}),\ \sigma_e^2)\ \sim \mathcal{N}(4.5, 3) \times \text{Exp}(0.02) \times \mathcal{N}(4.5, 3) \times \mathcal{U}(0, 1) \times \text{Exp}(0.02)$. In specifying the prior distribution, the main objective has been to use a weakly informed prior, thus, allowing the MCMC to explore a large volume of the parameter space without overwriting the signal in the data. This was verified by the nearly flat prior densities contrasting with sharply peaked posterior densities proving the presence of strong signal in the data (compare prior vs. posterior densities on supplementary fig. S8B, Supplementary Material online). To validate that the results were not sensitive to the parametrization and the definition of the prior, we tested other parametrizations and priors (e.g., $(\alpha,\ \theta,\ \sigma^2,\ \sigma_e^2) \sim \text{Exp}(0.01) \times \mathcal{U}(0, 100) \times \text{Exp}(0, 10^{-4}) \times \text{Exp}(0.01)$). These resulted in matching posterior means and HPDs for all sampled and derived parameters (not reported). The adaptive Metropolis MCMC was run for $4.2\text{E}+06$ iterations, of which the first $2\text{E}+05$ were used for warm-up and adaptation of the jump distribution variance–covariance matrix. The target acceptance rate was set to 0.01 and the thinning interval was set to 1,000. The convergence and mixing of the MCMC was validated by visual analysis (supplementary fig. S8A, Supplementary Material online) as well as by comparison to a parallel MCMC-chain started from a different initial state. Calculation of 95% HPD was done using the function "HPDinterval" from the coda package (Plummer et al. 2006).

Computer Simulations of the Toy Epidemiological Model

The parameters defining the within- and between-host dynamics used in the simulations are written in supplementary table S2, Supplementary Material online.

The simulations were implemented as stochastic random sampling of within- and between-host events (i.e., risky contact, transmission, mutation, diagnosis, death) in discrete time-steps of length 0.05 (arbitrary time-units). The transmission history as well as the history of within-host strain substitutions was preserved during the simulations in order to reproduce exact transmission trees and to extract donor and recipient values at moments of transmission for the calculation of b_0.

Software

This study relies on two accompanying R-packages:

- toyepidemic implementing the toy epidemiological model; available at https://github.com/venelin/toyepidemic.git, last accessed January 9, 2018; and
- patherit providing a common interface for evaluating the various heritability estimators on simulated and real data. The pair correlation and regression slope estimators are implemented as functions in this package; the phylogenetic heritability estimators (PMM and POUMM) are implemented as external calls to the R-package POUMM (Mitov and Stadler 2017). The patherit package is available at https://github.com/venelin/patherit.git, last accessed January 9, 2018.

External Dependencies

The following third-party R-packages were used: ape v3.4 (Paradis et al. 2004), data.table v1.9.6 (Dowle and Srinivasan 2017), adaptMCMC v1.1 (Scheidegger 2012), Rmpfr v0.6-0 (Maechler 2016), and coda v0.18-1 (Plummer et al. 2006). All programs have been run on R v3.2.4 (R Core Team 2016).

Data Availability

All scripts for performing the simulations and real data analyses presented in this paper are available at https://github.com/venelin/Estimating-Pathogen-Trait-Heritability.git, last accessed January 9, 2018. Large output data files from the toy model simulations are available upon request to the authors. The UK HIV data are not made available at the above address, because the authors do not have the right to redistribute this data (readers are referred to the UK drug resistance database).

Supplementary Material

Supplementary data are available at *Molecular Biology and Evolution* online.

Acknowledgments

References

Alizon S, von Wyl V, Stadler T, Kouyos RD, Yerly S, Hirschel B, Böni J, Shah C, Klimkait T, Furrer H. 2010. Phylogenetic approach reveals that virus genotype largely determines HIV set-point viral load. *PLoS Pathog.* 6(9):e1001123.

Anderson TJC, Williams JT, Nair S, Sudimack D, Barends M, Jaidee A, Price RN, Nosten F. 2010. Inferred relatedness and heritability in malaria parasites. *Proc R Soc B Biol Sci.* 277(1693):2531–2540.

Bachmann N, Turk T, Kadelka C, Marzel A, Shilaih M, Böni J, Aubert V, Klimkait T, Leventhal GE, Günthard HF, et al. 2017. Parent-offspring regression to estimate the heritability of an HIV-1 trait in a realistic setup. *Retrovirology* 14(1):33.

Bertels F, Marzel A, Leventhal G, Mitov V, Fellay J, Günthard HF, Böni J, Yerly S, Klimkait T, Aubert V, et al. 2018. Dissecting HIV virulence: heritability of setpoint viral load, CD4+ T cell decline and per-parasite pathogenicity. *Mol Biol Evol.* 35(1):27–37.

Bjorn-Mortensen K, Soborg B, Koch A, Ladefoged K, Merker M, Lillebaek T, Andersen AB, Niemann S, Kohl TA. 2016. Tracing *Mycobacterium tuberculosis* transmission by whole genome sequencing in a high incidence setting: a retrospective population-based study in East Greenland. *Sci Rep.* 6(1):33180.

Blanquart F, Wymant C, Cornelissen M, Gall A, Bakker M, Bezemer D, Hall M, Hillebregt M, Ong SH, Albert J, et al. 2017. Viral genetic variation accounts for a third of variability in HIV-1 set-point viral load in Europe. *Plos Biol.* 15(6):e2001855.

Bonhoeffer S, Fraser C, Leventhal GE. 2015. High heritability is compatible with the broad distribution of set point viral load in HIV carriers. *PLoS Pathog.* 11(2):e1004634–e1004634.

Dowle M, Srinivasan A 2017. data.table: Extension of 'data.frame'. R package version 1.10.4-3. https://CRAN.R-project.org/package=data.table.

Falconer DS, Mackay TFC. 1996. Introduction to quantitative genetics, 4th edition. Harlow, United Kingdom: Pearson Education Limited.

Felsenstein J. 1985. Phylogenies and the comparative method. *Am Nat.* 125(1):1–15.

Felsenstein J. 1988. Phylogenies and quantitative characters. *Annu Rev Ecol Syst.* 19(1):445–471.

FitzJohn RG. 2012. Diversitree: comparative phylogenetic analyses of diversification in R. *Methods Ecol Evol.* 3(6):1084–1092.

Fraser C, Hollingsworth TD, Chapman R, de Wolf F, Hanage WP. 2007. Variation in HIV-1 set-point viral load: epidemiological analysis and an evolutionary hypothesis. *Proc Natl Acad Sci U S A.* 104(44):17441–17446.

Fraser C, Lythgoe K, Leventhal GE, Shirreff G, Hollingsworth TD, Alizon S, Bonhoeffer S. 2014. Virulence and pathogenesis of HIV-1 infection: an evolutionary perspective. *Science* 343(6177):1243727–1243727.

Freckleton RP, Harvey PH, Pagel M. 2002. Phylogenetic analysis and comparative data: a test and review of evidence. *Am Nat.* 160(6):712–726.

Grimmett G, Stirzaker D. 2001. Probability and random processes. Oxford, UK: Oxford University Press.

Hansen TF. 1997. Stabilizing selection and the comparative analysis of adaptation. *Evolution Int J Org Evolution* 51(5):1341–1351.

Hansen TF, Bartoszek K. 2012. Interpreting the evolutionary regression: the interplay between observational and biological errors in phylogenetic comparative studies. *Syst Biol.* 61(3):413–425.

Hartl DL, Clark AG. 2007. Principles of population genetics. Sunderland, Massachusetts: Sinauer Associates.

Hecht FM, Hartogensis W, Bragg L, Bacchetti P, Atchison R, Grant R, Barbour J, Deeks SG. 2010. HIV RNA level in early infection is predicted by viral load in the transmission source. *AIDS* 24(7):941–945.

Hodcroft E, Hadfield JD, Fearnhill E, Phillips A, Dunn D, O'Shea S, Pillay D, Leigh Brown AJ. 2014. The contribution of viral genotype to plasma viral set-point in HIV infection. *PLoS Pathog.* 10(5):e1004112.

Hollingsworth TD, Laeyendecker O, Shirreff G, Donnelly CA, Serwadda D, Wawer MJ, Kiwanuka N, Nalugoda F, Collinson-Streng A, Ssempijja V, et al. 2010. HIV-1 transmitting couples have similar viral load set-points in Rakai, Uganda. *PLoS Pathog.* 6(5):e1000876.

Housworth EA, Martins EP, Lynch M. 2004. The phylogenetic mixed model. *Am Nat.* 163(1):84–96.

Keele BF, Giorgi EE, Salazar-Gonzalez JF, Decker JM, Pham KT, Salazar MG, Sun C, Grayson T, Wang S, Li H, et al. 2008. Identification and characterization of transmitted and early founder virus envelopes in primary HIV-1 infection. *Proc Natl Acad Sci U S A.* 105(21):7552–7557.

Keeling MJ, Rohani P. 2007. Modeling infectious diseases in humans and animals. Princeton, New Jersey: Princeton University Press.

Lande R. 1976. Natural-selection and random genetic drift in phenotypic evolution. *Evolution Int J Org Evolution* 30(2):314–334.

Leventhal GE, Bonhoeffer S. 2016. Potential pitfalls in estimating viral load heritability. *Trends Microbiol.* 24(9):687–698.

Lingappa JR, Thomas KK, Hughes JP, Baeten JM, Wald A, Farquhar C, de Bruyn G, Fife KH, Campbell MS, Kapiga S, et al. 2013. Partner characteristics predicting HIV-1 set point in sexually acquired HIV-1 among African seroconverters. *AIDS Res Hum Retroviruses.* 29(1):164–171.

Lynch M. 1991. Methods for the analysis of comparative data in evolutionary biology. *Evolution Int J Org Evolution* 45(5):1065–1080.

Lynch M, Walsh B. 1998. Genetics and analysis of quantitative traits. Sunderland, Massachusetts: Sinauer Associates Incorporated.

Lythgoe KA, Fraser C. 2012. New insights into the evolutionary rate of HIV-1 at the within-host and epidemiological levels. *Proc Biol Sci.* 279(1741):3367–3375.

Maechler M. 2016. Rmpfr: R MPFR - Multiple Precision Floating-Point Reliable. R package version 0.6-1. https://CRAN.R-project. org/package=Rmpfr

Mellors JW, Rinaldo CR, Gupta P, White RM, Todd JA, Kingsley LA. 1996. Prognosis in HIV-1 infection predicted by the quantity of virus in plasma. *Science* 272(5265):1167–1170.

Metzner KJ. 2016. HIV whole-genome sequencing now: answering still-open questions. *J Clin Microbiol.* 54(4):834–835.

Mitov V, Stadler T. 2016. The heritability of pathogen traits – definitions and estimators. bioRxiv 058503.

Mitov V, Stadler T. 2017. POUMM: An R-package for Bayesian Inference of Phylogenetic Heritability. bioRxiv 115089.

Paradis E, Claude J, Strimmer K. 2004. APE: analyses of phylogenetics and evolution in R language. *Bioinformatics* 20(2):289–290.

Plummer M, Best N, Cowles K, Vines K. 2006. CODA: convergence diagnosis and output analysis for MCMC. *R News* 6:7–11.

R Core Team. 2013. R: a language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing, https://www.R-project.org.

Scheidegger A. 2012. adaptMCMC: adaptive Monte Carlo Markov Chain sampler with coerced acceptance rate. R package version 1.1. https://CRAN.R-project.org/package=adaptMCMC.

Shirreff G, Alizon S, Cori A, Günthard HF, Laeyendecker O, van Sighem A, Bezemer D, Fraser C. 2013. How effectively can HIV phylogenies be used to measure heritability? *Evol Med Public Health* 2013(1):209–224.

Stearns SC, Koella JC. 2007. Evolution in health and disease. Oxford: OUP.

Tang J, Tang S, Lobashevsky E, Zulu I, Aldrovandi G, Allen S, Kaslow RA and Zambia-UAB HIV Research Project. 2004. HLA allele sharing and HIV type 1 viremia in seroconverting Zambians with known transmitting partners. *AIDS Res Hum Retroviruses* 20(1):19–25.

Uhlenbeck GE, Ornstein LS. 1930. On the theory of the Brownian motion. *Phys Rev.* 36(5):823–841.

van der Kuyl AC, Jurriaans S, Pollakis G, Bakker M, Cornelissen M. 2010. HIV RNA levels in transmission sources only weakly predict plasma viral load in recipients. *AIDS* 24(10):1607–1608.

Vihola M. 2012. Robust adaptive Metropolis algorithm with coerced acceptance rate. *Stat Comput.* 22(5):997–1008.

Yang Z. 2006. Computational molecular evolution. Oxford: OUP.

Yue L, Prentice HA, Farmer P, Song W, He D, Lakhi S, Goepfert P, Gilmour J, Allen S, Tang J, et al. 2013. Cumulative impact of host and viral factors on HIV-1 viral-load control during early infection. *J Virol.* 87(2):708–715.