

RESEARCH ARTICLE

High Heritability Is Compatible with the Broad Distribution of Set Point Viral Load in HIV Carriers

Sebastian Bonhoeffer^{1*}, Christophe Fraser², Gabriel E. Leventhal¹

1 Institute of Integrative Biology, ETH Zurich, Zurich, Switzerland, **2** Department of Infectious Disease Epidemiology, Imperial College London, London, United Kingdom

* sebastian.bonhoeffer@env.ethz.ch



 OPEN ACCESS

Citation: 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. doi:10.1371/journal.ppat.1004634

Editor: Ronald Swanstrom, University of North Carolina at Chapel Hill, UNITED STATES

Received: October 7, 2014

Accepted: December 16, 2014

Published: February 6, 2015

Copyright: © 2015 Bonhoeffer et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Funding: SB was supported in part by the European Research Council under the 7th Framework Programme of the European Commission (PBDR: Grant Agreement Number 268540). SB also received funding from the Swiss National Foundation (SNF). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

Abstract

Set point viral load in HIV patients ranges over several orders of magnitude and is a key determinant of disease progression in HIV. A number of recent studies have reported high heritability of set point viral load implying that viral genetic factors contribute substantially to the overall variation in viral load. The high heritability is surprising given the diversity of host factors associated with controlling viral infection. Here we develop an analytical model that describes the temporal changes of the distribution of set point viral load as a function of heritability. This model shows that high heritability is the most parsimonious explanation for the observed variance of set point viral load. Our results thus not only reinforce the credibility of previous estimates of heritability but also shed new light onto mechanisms of viral pathogenesis.

Author Summary

Following an initial peak in viremia, the viral load in HIV infected patients settles down to a set point which remains more or less stable during chronic HIV infection. This set point viral load is one of the key factors determining the rate of disease progression. The extent to which it is determined by the virus versus host genetics is thus central to developing a better understanding of disease progression. Here we develop an analytical model that describes the changes of the distribution of set point viral load in the HIV carrier population over a full cycle of transmission. Applying this model to patient data we find that the most parsimonious explanation for the observed large variation of set point viral load across HIV patients is that set point viral load is highly heritable from donors to recipients. This implies that set point viral load is to a considerable extent under the genetic control of the virus.

Introduction

The time course of viral load in HIV infected patients follows a characteristic pattern. During primary infection the viral load rapidly grows to very high levels. The peak viremia is attained within the first few weeks of infection. Thereafter the viral load declines rapidly over a period of several months and eventually settles down at a much lower level referred to as the viral set point. Set point viral load (spVL) is a central characteristic of the course of the disease. Firstly, the virus load measurements do fluctuate in patients, the time average of the viral load remains remarkably close to the spVL in most of patients over the time scale of several years [1, 2]. Secondly, higher spVL is associated with faster disease progression [3].

The stability of spVL within patients is in strong contrast to the enormous variation in spVL observed between patients. While variation in spVL between patients ranges over 3–4 orders of magnitude [3–6], the time trend over longitudinal viral load measurements typically changes by less than 0.1 log per year [1, 2]. Given that spVL is a key predictor of disease progression, there is considerable interest in identifying the host and viral genetic factors underlying the variation in spVL.

A well known example for the influence of naturally occurring variation in human genetic factors on viral load is the $\Delta 32$ deletion in the *CCR5* gene [7]. Moreover polymorphisms in HLA-B and C alleles have been associated with variance in virus load and genome-wide association studies (GWAS) showed that about 20% of the variance in log spVL can be attributed to specific single nucleotide polymorphisms [8–11]. 20% is likely a lower bound for the overall contribution of host genetic factors, because GWAS generally suffer from the problem that they can only identify common genetic variants with strong effects and do not account for epistatic effects between host genes [12].

Natural variation in the virus can also affect spVL. For example the transmission of a *nef*-deficient virus through a contaminated blood sample resulted in a low viral load in the recipients [13]. Moreover, several studies have reported a correlation between predicted replicative capacity and viral load [14–16]. As this prediction is based only on the viral genotype a patient carries, this implies that naturally occurring variation in viruses does affect viral load. A number of recent studies attempted to estimate the contribution of the viral genotype to the variation in spVL by quantifying the statistical association of viral load between donors and recipients either directly in donor-recipient pairs or through phylogenetic analysis [17–22]; for reviews see Müller et al. [23] and Fraser et al. [24]. A meta-analysis of previously published donor-recipient studies correcting for various co-factors such as age and sex yielded a heritability of 33% with a 95% confidence interval of 20–46% [24]. The two studies that inferred heritability based on phylogenetic methods provided the most extreme estimates with 5.7% reported by Hodcroft et al. [22] and 59% reported by Alizon et al. [21]. While the phylogenetic approaches have an advantage over the donor-recipient based approaches in that they can use much larger patient populations, it is currently unclear to what extent the underlying assumptions of the phylogenetic approaches of no selection and high frequency of sampling affect the robustness of these results.

The discrepant estimates call for a better quantitative understanding of the underlying factors determining heritability of log spVL in HIV. To this end we develop here a quantitative model that describes the change of the distribution of log spVL in a patient population in relation to heritability over a full transmission cycle. The model extends the approach of Shirreff et al. [25] and is similar in spirit to the integral projection models in ecology that are used to describe the temporal changes of distributions of a continuous phenotypic trait in populations [26–28]. In contrast to many applications in ecology, the application to distributions of log spVL has the advantage that all relevant processes and populations for which data are available,

are numerically well approximated by a Gaussian function. This fact enables us to obtain complete analytical understanding of how spVL changes through time based on a model parametrized by available data.

Results

We consider the change of the spVL distribution over one full reproduction cycle on the epidemiological level, i.e. from the current to the next generation of patients. We divide the patient population into “carriers” (HIV infected individuals prior to selection for transmission), “donors” (individuals that have been selected for transmission) and “recipients” (individuals that have just been infected by donors). Furthermore, we divide the reproduction cycle into three steps: (i) selection of donors from the carriers with replacement according to their transmission potential, (ii) transmission from donors to recipients, and (iii) intrahost evolution of the virus from the start of infection to the next transmission. Finally, we explicitly distinguish between factors contributing to set point viral load with regard to being transmissible (i.e. viral genetic factors) versus being non-transmissible (i.e. host genetic factors, environmental factors, or any interaction between host, virus, and the environment). A schematic overview over the effects of these steps on the distribution of log spVL is shown in [Fig. 1](#).

In the Supplementary Materials we show how the change of the spVL distribution can be computed for any distribution over a full transmission cycle. If all populations and processes are well approximated by Gaussian functions, then an approximation to the resulting log spVL distributions can be computed analytically (see [Methods](#) and Supplementary Materials). Assuming that the population is in equilibrium we obtain for the mean, \tilde{M}_C , variance, \tilde{V}_C , and heritability, h^2 , the following expressions:

$$\tilde{M}_C = \mu_o + \mu_i \left(1 + \frac{v_e + v_o}{\tilde{V}_C - v_e} \right), \tag{1}$$

$$\tilde{V}_C = \frac{v_t + v_i}{2} \left(1 + \sqrt{1 + 4 \frac{v_e + v_o}{v_t + v_i}} \right) + v_e, \tag{2}$$

$$h^2 = 1 - \frac{v_e}{\tilde{V}_C}. \tag{3}$$

Here the parameters μ_o and v_o characterize the transmission potential [6], i.e. the overall probability of a patient to transmit the infection as a function of log spVL ([Fig. 1\(o\)–\(i\)](#)). This transmission potential is given by the product of the rate of transmission per contact and the disease duration. As the former increases and the latter decrease with increasing spVL, the transmission potential has a maximum at intermediate levels of spVL [6]. The parameter v_e gives the variance of the contribution of host/environmental effects on log spVL. The parameter v_t describes the variance due to the bottleneck at transmission from donor to recipient, as a founder strain is selected randomly from the diverse population in the donor ([Fig. 1\(i\)–\(ii\)](#)). The parameters μ_i and v_i describe the mean and variance of the contribution of intrahost evolution to log spVL ([Fig. 1\(ii\)–\(iii\)](#)).

Our model assumes that the bottleneck at transmission is neutral with regard to selection on set point viral load. Note, that the assumption is without loss of generality. This is important because there is evidence for selection at transmission [29], although it is unclear whether selection acts on spVL. Any selective effect at transmission, however, can be subsumed into the parameter μ_i . Hence, the effect of selection is effectively incorporated in our model.

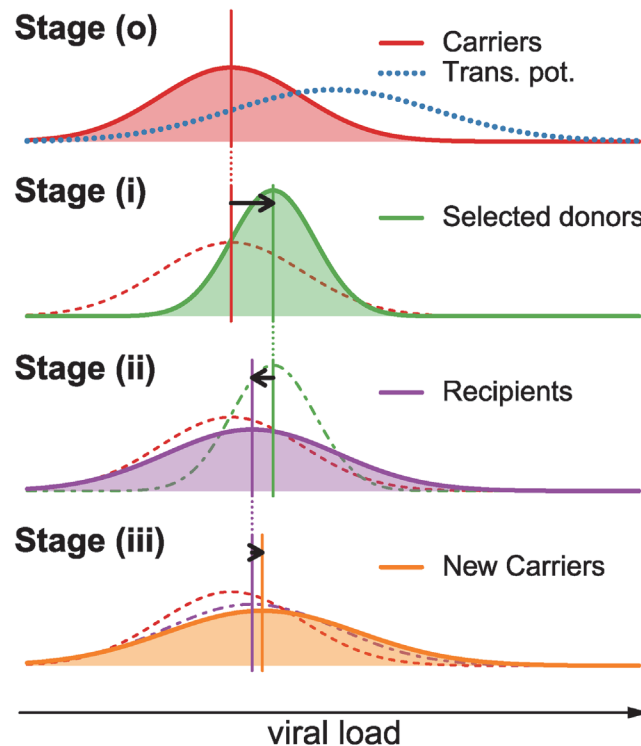


Fig 1. Graphical example of the change in the distribution of log spVL in the population over one reproduction cycle. During one full reproduction cycle, the distribution goes through the following steps: (o) The log spVL distribution within a population follow a Gaussian function with mean M_C and variance V_C (red curve). The transmission potential (blue dotted line) selects a subset of this population as donors (see Equation 5). (i) The transmission potential selects donors from the carrier population in (o) with mean M_D that lies between the mean of the carriers, M_C , and the mean of the transmission potential, μ_o . The resulting variance in log spVL in the selected donors is smaller than in the carrier population (see Equation 7). (ii) The selected donors transmit to new hosts, thus randomizing the host/environmental contributions and lowering the population mean log spVL and increasing the population variance. The variance is further increased by a transmission bottleneck and sampling effect on the level of the individual donors. (iii) Within-host evolution of log spVL may further increase or decrease the population mean, while always increasing or not affecting the variance. This completes a full reproduction cycle. In equilibrium, the individual changes in mean and variance in stages (i), (ii) and (iii) is such that the overall change in mean and variance from stage (o) to (iii) is zero.

doi:10.1371/journal.ppat.1004634.g001

The parameter for the mean contribution by the host/environment, μ_e , does not appear in equations 1 or 2. This is because the equations refer to the phenotypic value of spVL, i.e. the sum of the genetic contributions of the virus and the contributions from the host/environment. Any large environmental/host effect on the mean can always be compensated by correspondingly strong genetic effect of the virus on the mean but with opposite sign.

The above results are applicable if, (i) if the population is approximately in equilibrium, and (ii) all populations and processes are numerically well approximated by Gaussian functions.

Assumption (i) has been discussed in detail previously [6, 25, 30, 31]. In essence, this assumption is supported by three observations. Firstly, the mean of the spVL distribution coincides with the optimum of the transmission potential (see Fig. 2 and Fraser et al. [6]). Secondly, the rate of change of spVL has decreased over the last 25 years [31]. Thirdly, the rate of evolution is sufficiently rapid such that a spVL that is optimal for transmission could have evolved over the course of the epidemic [25]. These findings suggest that the distribution of set point

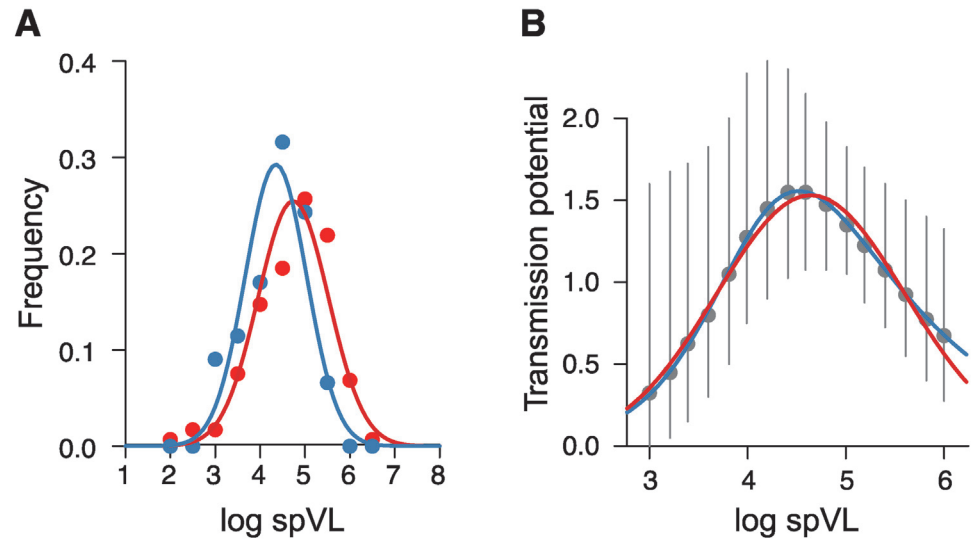


Fig 2. Viral load distributions and transmission potential estimated from patient cohorts as extracted from the corresponding graphs in Fraser et al. [6]. (A) The viral load distributions in the a Zambian and a Dutch cohort (see Fraser et al. [6]). The lines correspond to the best fits of a Gaussian to the distribution of log spVL. The null hypothesis that log spVL is normally distributed cannot be rejected based on a test that tests whether the residuals between model and fit themselves are normally distributed. The estimated mean and standard deviation are 4.74 and 0.61 for the Zambian data (red) and 4.35 and 0.47 for the Amsterdam data (blue). (B) The transmission probability according to the functions for transmissibility and duration of disease as a function of viral load as provided in Fraser et al. [6]. The grey circles and lines represent the mean and 95% confidence interval of the transmission potential as estimated from the Zambian and Amsterdam cohorts by Fraser et al. [6]. The blue line represents the corresponding theoretically derived transmission potential as provided by Fraser et al. [6]. The red line corresponds to the best fit of a log Gaussian to the estimated transmission potential. The parameters of the fitted log Gaussian are $\mu_o = 4.64 \pm 0.021$ and $v_o = 0.96 \pm 0.025$ (estimate \pm standard deviation).

doi:10.1371/journal.ppat.1004634.g002

viral load is indeed approximately in equilibrium, which in turn makes it is plausible to assume that the environmental and genetic factors determining set point viral load are also in equilibrium.

Regarding assumption (ii), we note that a Gaussian function describes a distribution or process by a main effect (mean) and some variational noise (variance). Thus in absence of any better knowledge, a Gaussian distribution is a natural starting point to describe any process and simply represents a second order approximation to an unknown distribution. We can assess the validity of describing the distributions of spVL in carriers and the transmission potential graphically using available data. Inspection of Fig. 2A and Figure S1 in Supplementary S1 Text shows that the viral load amongst carriers is indeed numerically well approximated by a Gaussian with mean log spVL, $M_C \approx 4.5$, and variance in log spVL, $V_C \approx 0.5$. Also the fit of a Gaussian to the transmission potential (see Fig. 2B) is a very good approximation (mean $\mu_o \approx 4.6$ and variance $v_o \approx 1.0$), even though the transmission potential as estimated by Fraser et al. [6] is slightly right-skewed.

There are no data to inform the shape of the processes of transmission and intrahost evolution. Using a description that has a mean effect with some variation around this mean is natural. Nonetheless, we test the effect of numerical deviations from a Gaussian with the following simulations. Firstly, we use the exact right-skewed transmission potential as given by Fraser et al. [6]. The analytical approximations for the distribution of the population in equilibrium remain excellent when the substantial deviations of the transmission potential from a Gaussian

are incorporated (see Figure S2). Secondly, we study the robustness towards deviations from Gaussian functions in the processes describing intrahost evolution and the transmission bottleneck. Even when both processes are strongly skewed, the analytical approximations for mean and variance are excellent (typically less than 2% deviation, see Figure S3 in Supplementary S1 Text).

To assess what heritability values are compatible with the observed mean and variance of log SPVL in the carrier population we take a simple approach that is in essence Approximate Bayesian Computing with rejection sampling. To this end we define plausible prior distributions for the parameters of the model. Sampling randomly from the priors we determine the resulting means and variances of log spVL in carriers and reject sets of parameters that lead to means and variances outside a defined permissible range. The set of accepted parameters gives the posterior distribution.

For the range of permissible mean log spVL we assume $4 < \tilde{M}_C < 5$, which is compatible but somewhat larger than the observed range in the studies reported by Fraser et al. [6] and Gekus et al. [5] (see Fig. 2A and Supplementary Materials, Section E). For the permissible range of variances of log spVL we assume that $0.3 < \tilde{V}_C < 0.8$, which again is compatible but somewhat larger than the values reported by Fraser et al. [6] and Gekus et al. [5] (see Supplementary Materials, Section E).

We use uniform priors for all parameters. The parameters μ_o and ν_o , which describe mean and variance of the transmission potential, have thus far only been estimated only by a single peer reviewed study (Fraser et al. [6] and Fig. 2B; see also [32]). To account for uncertainty in the estimates of these parameters we use $4 < \mu_o < 5$ and $0.5 < \nu_o < 1.5$. Estimates for remaining parameters cannot be easily derived from the existing literature. To account for uncertainty in these parameters we assume $0 < \nu_e < 1$; $-1 < \mu_i < 1$; $0 < \nu_i < 0.3$ and $0 < \nu_t < 0.3$.

Fig. 3 shows the posterior parameter distribution from the rejection sampling. Different colors in the scatter plots indicate different levels of mean heritability at given parameter combinations. The contour lines show the density of posterior distribution. The key result shown in

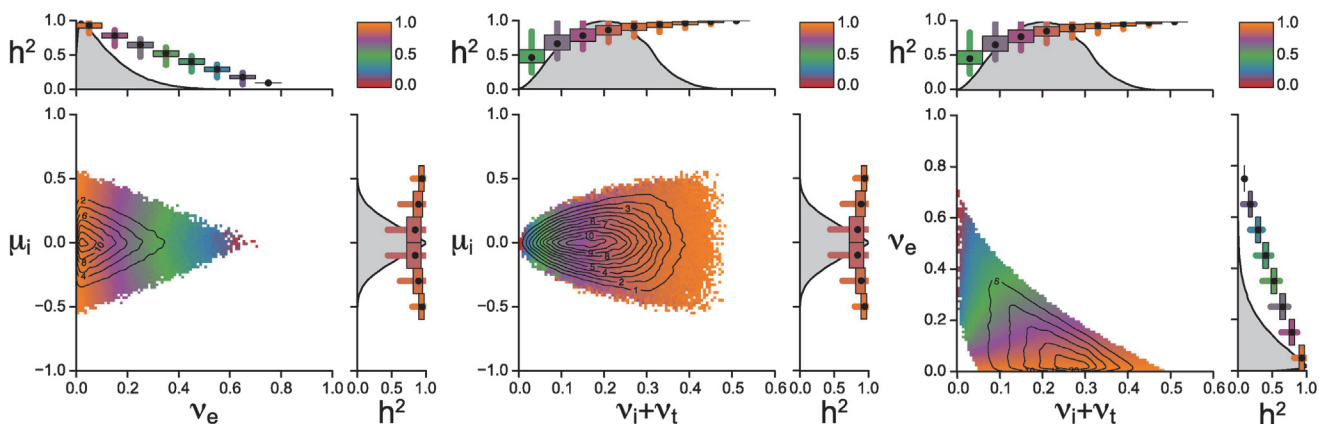


Fig 3. Posterior distribution of parameters from the rejection sampler. We report pairwise scatterplots of the parameters μ_i , ν_e and the compound parameter $\nu_i + \nu_t$, since these two parameters only appear as a sum in all equations for the mean, variance and heritability. 10^7 random sets of parameter values are sampled randomly from the uniform priors described by the ranges on the x and y axes. Around 1.9% of the randomly generated parameter combinations yield values for mean and variance of log spVL that are compatible with the acceptance criterion $4 < M_C < 5$, $0.3 < V_C < 0.8$. The contour lines show the two-dimensional kernel density estimate of the posterior sample. The colours reflect the mean heritability of binned parameter combinations and are stacked such that points with lower heritability lie on top of points with higher heritability. The small plots to the top and right of the scatterplots show the posterior density estimate along a single parameter dimension, as well as the mean (black dot), and 50% (boxes) and 95% quantiles (lines) of heritabilities along those parameter dimensions. Most of the probability mass occurs at low values of μ_i and ν_e .

doi:10.1371/journal.ppat.1004634.g003

the figure is that the majority of accepted parameter values result in high values of heritability (purple to orange color at contour lines of highest posterior densities). While low values of heritability are also compatible with the observed mean and variance of log spVL, they occur rarely in the posterior sample and are at the edges of the prior distributions (red to blue areas). The center of mass of the posterior sample is in areas with high heritability, higher in fact than what would seem compatible with current estimates of heritability and host genetic factors. There are two factors not included in this analysis: measurement error of spVL and prior knowledge of the host contribution to spVL. Increasing the measurement accuracy of spVL would increase heritability estimates based on both donor-recipient pairs and phylogenetic inference. Incorporating prior knowledge of the host genetic contribution would set an upper bound on the estimates of heritability in our analysis. Thus accounting for these two factors bring the center of mass of the heritability distribution closer to the measured values of heritability.

The figure also highlights that generally wider priors would not change the posterior distribution because parameter values at the upper end of priors are never accepted. The change of the mean capacity of the virus to induce spVL through intrahost evolution, μ_i , is restricted to values smaller than 0.6 and decreases with increasing variance generated by the host/environment effects, v_e . Increasing v_e corresponds to decreasing heritability (see [eq. 3](#)) and thus high levels of μ_i require high levels of heritability. The center of mass of the posterior sample suggests that the most parsimonious explanation of the observed mean and variance of log spVL implies both small intrahost evolution and high heritability.

One criticism leveled against the transmission potential as quantified in [\[6\]](#) is that it does not appropriately reflect transmissions occurring during the acute or the AIDS phase. In the Supplementary Material, Section F.3, we show that our quantitative results are robust towards using a corrected transmission potential.

Discussion

The above analysis shows that the most parsimonious explanation of the observed distribution of spVL in HIV carrier populations requires high heritability of spVL. Although low heritability values are also compatible with the observed distribution of spVL in HIV carrier populations, parameter combinations resulting in these low values have a small probability and occur at the edge of the realistic parameter range. The skepticism with which the estimated heritability values have been met in the field suggests that the general expectation is that heritability of spVL should be low. In contrast, our analysis shows that high heritability values are not only compatible with, but are also the more parsimonious explanation of the observed distribution in spVL in HIV carrier populations.

Low heritability only occurs if the processes of intrahost evolution and the transmission bottleneck have a weak effect on spVL, i.e. if the parameters μ_i , v_i and v_t are small. An intuition can be obtained by noting that in equilibrium the variance generating and variance eliminating processes balance out. The transmission potential only exerts weak selection on log spVL and therefore only marginally reduces variance. The decrease of variance by selection for transmission has to be compensated by an increase in variance by intrahost evolution and the transmission bottleneck. For too low heritability, the genetic variance generated by intrahost evolution and transmission bottlenecks would overwhelm the reduction of variance due to selection by the transmission potential. While there are to our knowledge no data that allow to estimate the variance generated at transmission, v_t , the posterior distributions of μ_i and v_i are broadly compatible with the observed changes of virus load within patients [\[1, 2, 33\]](#).

Taken together our analysis suggests that the most parsimonious explanation of the distribution of log spVL is high h^2 but low v_t , μ_i and v_i . Hence, heritability is high while the processes

of intrahost evolution and transmission bottleneck have a small effect on the capacity of the virus to modulate log spVL. High heritability implies a substantial genetic control of the log spVL by the virus. The observation that at the same time the contribution of intrahost evolution to spVL is small raises an interesting question: How can a strongly heritable trait show little intrahost evolution? Given the otherwise ample evidence for rapid intrahost evolution of HIV such as escape from drugs or the immune response, the absence of intrahost evolution of spVL is surprising. Generally a trait is expected to respond to selection, if (i) the trait is heritable, (ii) there is phenotypic variation of the trait in a population, and (iii) the trait is linked to fitness. That spVL is heritable has been reported previously [23, 24] and our analysis reinforces the credibility of these findings. That there is phenotypic variation in the control of spVL by the virus is plausible given the large genetic variation of the virus population within an individual. What remains is whether it is conceivable that the capacity of a viral genotype to induce spVL is only weakly linked to fitness. One hypothesis that could reconcile high heritability with little intrahost evolution is that variation in viral load between patients is in part due to virus-induced activation of target cells. Difference in activation rate of target cells has previously been argued to account for a substantial part of the variation in viral load [4]. Furthermore, if target cell activation is at least partially under the control of the virus, then this control may indeed be weakly linked to intrahost fitness. If the target cell activation is systemic (i.e. not locally confined to the inducing virus) then increased target cell activation increases the pool of susceptible cells, but the benefit of increased target cell activation is not confined to the producer virus. As a result selection for virus induced activation rate is expected to be neutral or nearly neutral [34]. Indeed, an explicit model of the evolution of log spVL for a virus induced control of target cell activation can reconcile high heritability with absence of intrahost evolution [35].

Our modeling approach is based on describing how the distribution of a continuous phenotypic trait, here log spVL, changes in a population over a full cycle of reproduction. This approach is closely related to the method of integral projection models, which has been developed and widely applied in ecology and population biology [26–28, 36, 37]. The approach can in principle describe how arbitrary distributions change over time as a function of processes such as selection and reproduction. Here we are able to obtain a full analytical description of the temporal change of the spVL distribution, because all relevant distributions and processes can be well approximated by Gaussian functions. We also show that our analytical results remain robust even for substantial deviations numerical deviations from Gaussian functions (see Supplementary Materials, Section F). Moreover, the model can be parametrized on the basis of available data. There are ample data for mean and variance of spVL and also most of the parameters can be confined to plausible ranges based on the literature.

Our study clearly supports that high heritability is compatible with the observed distribution of log spVL in HIV carriers. High heritability of spVL does not preclude that also the host genotype has a considerable effect on virus load. However, it does lead to the expectation that over the course of infection the capacity to induce higher spVL should increase considerably unless this capacity is only weakly linked to intrahost fitness. This sheds new light onto the mechanisms controlling viral load. There should be identifiable genetic variation in the virus population that is associated with viral load, and moreover, the loci associated with control of viral load should be weakly linked to intrahost fitness. Genome-wide association studies mapping viral genetic polymorphisms to variance in log spVL seem a natural approach to test this prediction. A recent study by Bartha et al. [38] was unable to identify any statistical associations, but was powered only to detect individual non-synonymous mutations with an effect size of >4% on heritability. Larger studies will thus be required to identify whether and which viral polymorphisms are associated with set point viral load.

Methods

In the following sections we derive an analytical model that describes the change of mean and variance of spVL in the population of HIV carriers as a function of the heritability of spVL. We account for the virus and host effects by subdividing the phenotype (i.e. log spVL) into genetic and environmental/host components. Generally we denote the changes in mean and variance in the carrier, donor, and recipient populations with the subscripts C , D and R , respectively. We use greek letters for the parameters of the model and latin letters for the variables. When referring to the phenotype (i.e. log spVL) we use upper-case letters and when referring to the genotype we use lower-case letters.

Distribution of log spVL in carrier population

The spVL in a patient is generally determined by viral genetic factors, host genetic factors, the environment and interactions between these factors. Since only the virus is transmitted from donors to recipients, we subsume all non-transmissible effects such as the host genetic factors, environmental effects and all interactions between host, virus and the environment generically under “environmental effects”, e . The transmissible effects due to the viral genotype are the “genotypic effects”, g . The “phenotype” spVL is then given by $g+e$.

We assume that the distribution of log spVL in the carrier population is given by a normal distribution $\mathcal{N}(M_C, V_C)$, where M_C and V_C are the mean and variance, respectively. The transmission potential, defined as the overall probability of transmission of an HIV carrier integrated over the entire course of the disease, is assumed to be a function of log spVL which can be well approximated by a normal distribution $\mathcal{N}(\mu_o, \nu_o)$ (see Fig. 2). Here μ_o is the log spVL at which the transmission potential is maximal and ν_o characterizes how strongly the transmission potential selects for transmission at μ_o .

We assume that g and e are independent and normally distributed in the carrier population with $\mathcal{N}(m_C, v_C)$ and $\mathcal{N}(\mu_e, \nu_e)$, respectively. Here m_C and v_C are the variables that describe the mean and variance of the distribution of viral genotypes in the carrier population. Note that here the independence of g and e refers to the quantitative contribution of virus and host to spVL. Importantly, this independence does not imply an absence of virus genotype by host genotype interactions, such as an interaction between a particular viral epitope and a host HLA molecule. Genotype by genotype interactions are non-transmissible and thus subsumed in e . The parameters μ_e and ν_e describe mean and variance of the distribution of environmental effects, which comprise host effects, interactions and any non-transmissible effect. The distribution of phenotype log spVL in the carrier population is then given by a normal distribution with mean and variance,

$$M_C = m_C + \mu_e, \quad \text{and} \quad V_C = v_C + \nu_e. \quad (4)$$

Selection of donors

Selection for transmission acts on log spVL, i. e. on the sum of the genotypic and environmental effects, and is given by the transmission potential. Specifically, the probability of transmission for a given log spVL, ϕ , is given by (see Fig. 2B),

$$S(\phi) = \frac{1}{\sqrt{2\pi\nu_o}} e^{-\frac{(\phi-\mu_o)^2}{2\nu_o}}. \quad (5)$$

Applying the above transmission potential to the carrier population, we find that the genotype and phenotype in the donor population are again normally distributed (see Supplementary Materials, Equations B6 and B7). The donor genotype has mean and variance,

$$m_D = \frac{m_C(v_e + v_o) + (\mu_o - \mu_e)v_C}{v_C + v_e + v_o}, \quad \text{and} \quad v_D = \frac{v_C(v_e + v_o)}{v_e + v_o + v_C}. \quad (6)$$

The donor phenotype has mean and variance (see Supplementary Materials, Equations B8 and B9),

$$M_D = \frac{M_C v_o + \mu_o V_C}{v_o + V_C}, \quad \text{and} \quad V_D = \frac{V_C v_o}{v_o + V_C}. \quad (7)$$

Note, that the mean and variance of the environmental effects (i.e. the host effect) is not given by the differences between the phenotypic and genotypic values, because environment and genotype in the donors are correlated. This is because selection for transmission acts on the sum of environmental and genotypic effects. In other words selection for transmission selects a subset of viral genotypes and host genotypes, and host and viral genotypes are correlated, because selection operates on their combined effect.

Transmission to recipients

When the virus is transmitted from the donor to the recipient population, the virus is “harvested” from a non-random distribution of environmental effects (and thus also from a non-random set of hosts). The harvested virus is then redistributed over a random set of new hosts/ environments in the recipient population. Thus all environmental effects in the donor population are erased at transmission and the environmental contribution in the recipients is redrawn from $\mathcal{N}(\mu_e, v_e)$. To account for the fact that the virus population experiences a strong bottleneck from recipient to donor, we assume that the viral genotype is not transmitted exactly from donor to recipient but instead is assumed to be randomly drawn out of a distribution of genotypes in the donor patient. Assuming that this distribution is normal with mean m_D and variance v_t we obtain that both genotype and phenotype in the donor population are normally distributed. The recipient genotype has mean and variance,

$$m_R = m_D, \quad \text{and} \quad v_R = v_D + v_t. \quad (8)$$

The recipient phenotype has mean and variance,

$$M_R = m_R + \mu_e^0, \quad \text{and} \quad V_R = v_R + v_e^0, \quad (9)$$

where μ_e^0 and v_e^0 are the mean and variance of the host/environmental effects prior to infection. The environmental effects are redrawn randomly, because they are not inherited from one transmission to the next. Note, that we assume here that the bottleneck at transmission is neutral. This assumption does not imply that there is no selection at the transmission stage, but rather that the bottleneck is neutral with regard to the spVL that the transmitted strains will eventually cause. Any selection at and after transmission on the viral genotypic contribution to log spVL is subsumed in the next step, intrahost evolution.

Intrahost evolution

After transmission the virus population in the recipient may change in a directed fashion according to intrahost evolution. Assuming that the overall change of the viral genotype due to

intra-host evolution can be approximated by a normal distribution we find that the distribution of genotypes and phenotypes in the next generation of carriers, C' is again normal. The distribution of the genotypes has a mean and variance

$$m_{C'} = m_R + \mu_i, \quad \text{and} \quad v_{C'} = v_R + v_i. \quad (10)$$

The parameter μ_i thus describes any genetic change in the virus that affects log spVL across all patients in the same way. The parameter v_i describes genetic changes that affect log spVL in a manner that is specific to the patient, i.e. it describes the effect of changes of log spVL due to genetic interactions between virus and host. As the environmental effects comprise the immune response by the host, the mean and variance in environmental effects may change in co-evolution with the virus through μ_e^i and v_e^i , respectively. Thus we obtain for mean and variance of the distribution of phenotypes,

$$M_{C'} = m_{C'} + \mu_e^0 + \mu_e^i, \quad \text{and} \quad V_{C'} = v_{C'} + v_e^0 + v_e^i. \quad (11)$$

Note, that any selection for spVL at the transmission bottleneck can now be interpreted as a genotypic change that occurs during intra-host evolution. Thus the overall model is appropriate both for non-selective and selective bottlenecks.

Heritability

Heritability, h^2 , is defined as fraction of genotypic variance relative to phenotypic variance in the carrier population [39]. Thus we have,

$$h^2 = \frac{v_C}{V_C} = 1 - \frac{v_e}{V_C}. \quad (12)$$

Heritability can be estimated in a parent-offspring regression [39], where h^2 is equal to the regression slope b . Donor-recipient pairs can be seen as parent-offspring pairs, where care must be taken since the donors are not randomly selected from the carrier population but are selected according to the transmission potential. Since, however, we are measuring the heritability of spVL and donors are selected based on spVL, the regression of recipients on selected donors is equal to heritability of spVL in carriers [24, 39].

Mean and variance of log spVL at equilibrium

We now have a complete analytical description how mean and variance of log spVL change from the current to the next generation of carriers. The fact that the log spVL that maximizes the transmission potential and the mean of the distribution of log spVL in the carrier populations (see Fig. 2 and Fraser et al. [6]) are both around 4.5, we can assume that the process is roughly at equilibrium. In equilibrium we have that the mean and variance of the distribution of phenotypes does not change, i.e. $M_{C'} = M_C$ and $V_{C'} = V_C$. This will be fulfilled if the genetic and environmental contributions are also at equilibrium, implying in particular that $\mu_e = \mu_e^0 + \mu_e^i$ and $v_e = v_e^0 + v_e^i$ (see Supplementary Materials section C.1). Using Equation 12 we can express the equilibrium mean and variance of log spVL as a function of v_e , the variance of the

contribution of the host/environment to log spVL (see Supplementary Materials, Equations C9 and C10),

$$\tilde{M}_C = \mu_o + \mu_i \left(1 + \frac{v_e + v_o}{\tilde{V}_C - v_e} \right), \quad (13)$$

$$\tilde{V}_C = \frac{v_t + v_i}{2} \left(1 + \sqrt{1 + 4 \frac{v_e + v_o}{v_t + v_i}} \right) + v_e. \quad (14)$$

or as a function the heritability h^2 (see Supplementary Materials, Equations C12 and C13),

$$\tilde{M}_C = \mu_o + \mu_i \left(1 + \frac{(1 - h^2) \tilde{V}_C + v_o}{h^2 \tilde{V}_C} \right), \quad (15)$$

$$\tilde{V}_C = \frac{v_t + v_i}{2(h^2)^2} \left(1 + \sqrt{1 + \frac{4(h^2)^2 v_o}{v_t + v_i}} \right). \quad (16)$$

Supporting Information

S1 Text. Supplementary Methods and Figures.

(PDF)

Acknowledgments

We thank Helen Alexander, Roland Regös, and Viktor Müller for helpful and stimulating discussions.

Author Contributions

Conceived and designed the experiments: SB GEL CF. Performed the experiments: SB GEL CF. Analyzed the data: SB GEL CF. Contributed reagents/materials/analysis tools: SB GEL CF. Wrote the paper: SB GEL CF.

References

1. O'Brien TR, Rosenberg PS, Yellin F, Goedert JJ (1998) Longitudinal HIV-1 RNA levels in a cohort of homosexual men. *J Acquir Immune Defic Syndr Hum Retrovirol* 18: 155–61. doi: [10.1097/00042560-199806010-00007](https://doi.org/10.1097/00042560-199806010-00007) PMID: [9637580](https://pubmed.ncbi.nlm.nih.gov/9637580/)
2. Sabin CA, Devereux H, Phillips AN, Hill A, Janossy G, et al. (2000) Course of viral load throughout HIV-1 infection. *J Acquir Immune Defic Syndr* 23: 172–7. doi: [10.1097/00042560-200002010-00009](https://doi.org/10.1097/00042560-200002010-00009) PMID: [14607066](https://pubmed.ncbi.nlm.nih.gov/14607066/)
3. Mellors JW, Rinaldo CR Jr, Gupta P, White RM, Todd JA, et al. (1996) Prognosis in HIV-1 infection predicted by the quantity of virus in plasma. *Science* 272: 1167–70. doi: [10.1126/science.272.5265.1167](https://doi.org/10.1126/science.272.5265.1167) PMID: [8638160](https://pubmed.ncbi.nlm.nih.gov/8638160/)
4. Bonhoeffer S, Funk GA, Günthard HF, Fischer M, Müller V (2003) Glancing behind virus load variation in HIV-1 infection. *Trends Microbiol* 11: 499–504. doi: [10.1016/j.tim.2003.09.002](https://doi.org/10.1016/j.tim.2003.09.002) PMID: [14607066](https://pubmed.ncbi.nlm.nih.gov/14607066/)
5. Geskus RB, Prins M, Hubert JB, Miedema F, Berkhout B, et al. (2007) The HIV RNA setpoint theory revisited. *Retrovirology* 4: 65. doi: [10.1186/1742-4690-4-65](https://doi.org/10.1186/1742-4690-4-65) PMID: [17888148](https://pubmed.ncbi.nlm.nih.gov/17888148/)
6. 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: 17441–6. doi: [10.1073/pnas.0708559104](https://doi.org/10.1073/pnas.0708559104) PMID: [17954909](https://pubmed.ncbi.nlm.nih.gov/17954909/)

7. Meyer L, Magierowska M, Hubert JB, Rouzioux C, Deveau C, et al. (1997) Early protective effect of CCR-5 delta 32 heterozygosity on HIV-1 disease progression: relationship with viral load. The SEROCO Study Group. *AIDS* 11: F73–8. doi: [10.1097/00002030-199711000-00001](https://doi.org/10.1097/00002030-199711000-00001) PMID: [9302436](https://pubmed.ncbi.nlm.nih.gov/9302436/)
8. Fellay J, Shianna KV, Ge D, Colombo S, Ledergerber B, et al. (2007) A whole-genome association study of major determinants for host control of HIV-1. *Science* 317: 944–7. doi: [10.1126/science.1143767](https://doi.org/10.1126/science.1143767) PMID: [17641165](https://pubmed.ncbi.nlm.nih.gov/17641165/)
9. Fellay J, Ge D, Shianna KV, Colombo S, Ledergerber B, et al. (2009) Common genetic variation and the control of HIV-1 in humans. *PLoS Genet* 5: e1000791. doi: [10.1371/journal.pgen.1000791](https://doi.org/10.1371/journal.pgen.1000791) PMID: [20041166](https://pubmed.ncbi.nlm.nih.gov/20041166/)
10. Dalmaso C, Carpentier W, Meyer L, Rouzioux C, Goujard C, et al. (2008) Distinct genetic loci control plasma HIV-RNA and cellular HIV-DNA levels in HIV-1 infection: the ANRS Genome Wide Association O1 study. *PLoS One* 3: e3907. doi: [10.1371/journal.pone.0003907](https://doi.org/10.1371/journal.pone.0003907) PMID: [19107206](https://pubmed.ncbi.nlm.nih.gov/19107206/)
11. Limou S, Le Clerc S, Coulonges C, Carpentier W, Dina C, et al. (2009) Genomewide association study of an AIDS-nonprogression cohort emphasizes the role played by HLA genes (ANRS Genomewide Association Study 02). *J Infect Dis* 199: 419–26. doi: [10.1086/596067](https://doi.org/10.1086/596067) PMID: [19115949](https://pubmed.ncbi.nlm.nih.gov/19115949/)
12. Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorf LA, et al. (2009) Finding the missing heritability of complex diseases. *Nature* 461: 747–53. doi: [10.1038/nature08494](https://doi.org/10.1038/nature08494) PMID: [19812666](https://pubmed.ncbi.nlm.nih.gov/19812666/)
13. Learmont JC, Geczy AF, Mills J, Ashton LJ, Raynes-Greenow CH, et al. (1999) Immunologic and virologic status after 14 to 18 years of infection with an attenuated strain of HIV-1. A report from the Sydney Blood Bank Cohort. *N Engl J Med* 340: 1715–22. doi: [10.1056/NEJM199906033402203](https://doi.org/10.1056/NEJM199906033402203) PMID: [10352163](https://pubmed.ncbi.nlm.nih.gov/10352163/)
14. Kouyos RD, von Wyl V, Hinkley T, Petropoulos CJ, Haddad M, et al. (2011) Assessing predicted HIV-1 replicative capacity in a clinical setting. *PLoS Pathog* 7: e1002321. doi: [10.1371/journal.ppat.1002321](https://doi.org/10.1371/journal.ppat.1002321) PMID: [22072960](https://pubmed.ncbi.nlm.nih.gov/22072960/)
15. Quiñones-Mateu ME, Ball SC, Marozsan AJ, Torre VS, Albright JL, et al. (2000) A dual infection/competition assay shows a correlation between ex vivo human immunodeficiency virus type 1 fitness and disease progression. *J Virol* 74: 9222–33. doi: [10.1128/JVI.74.19.9222-9233.2000](https://doi.org/10.1128/JVI.74.19.9222-9233.2000) PMID: [10982369](https://pubmed.ncbi.nlm.nih.gov/10982369/)
16. Barbour JD, Hecht FM, Wrin T, Segal MR, Ramstead CA, et al. (2004) Higher CD4+ T cell counts associated with low viral pol replication capacity among treatment-naïve adults in early HIV-1 infection. *J Infect Dis* 190: 251–6. doi: [10.1086/422036](https://doi.org/10.1086/422036) PMID: [15216458](https://pubmed.ncbi.nlm.nih.gov/15216458/)
17. Tang J, Tang S, Lobashevsky E, Zulu I, Aldrovandi G, et al. (2004) HLA allele sharing and HIV type 1 viremia in seroconverting Zambians with known transmitting partners. *AIDS Res Hum Retroviruses* 20: 19–25. doi: [10.1089/088922204322749468](https://doi.org/10.1089/088922204322749468) PMID: [15000695](https://pubmed.ncbi.nlm.nih.gov/15000695/)
18. Hecht FM, Hartogensis W, Bragg L, Bacchetti P, Atchison R, et al. (2010) HIV RNA level in early infection is predicted by viral load in the transmission source. *AIDS* 24: 941–5. doi: [10.1097/QAD.0b013e328337b12e](https://doi.org/10.1097/QAD.0b013e328337b12e) PMID: [20168202](https://pubmed.ncbi.nlm.nih.gov/20168202/)
19. Hollingsworth TD, Laeyendecker O, Shirreff G, Donnelly CA, Serwadda D, et al. (2010) HIV-1 transmitting couples have similar viral load set-points in Rakai, Uganda. *PLoS Pathog* 6: e1000876. doi: [10.1371/journal.ppat.1000876](https://doi.org/10.1371/journal.ppat.1000876) PMID: [20463808](https://pubmed.ncbi.nlm.nih.gov/20463808/)
20. 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: 1607–8. doi: [10.1097/QAD.0b013e32833b318f](https://doi.org/10.1097/QAD.0b013e32833b318f) PMID: [20539098](https://pubmed.ncbi.nlm.nih.gov/20539098/)
21. Alizon S, von Wyl V, Stadler T, Kouyos RD, Yerly S, et al. (2010) Phylogenetic approach reveals that virus genotype largely determines HIV set-point viral load. *PLoS Pathog* 6: e1001123. doi: [10.1371/journal.ppat.1001123](https://doi.org/10.1371/journal.ppat.1001123) PMID: [20941398](https://pubmed.ncbi.nlm.nih.gov/20941398/)
22. Hodcroft E, Hadfield JD, Fearnhill E, Phillips A, Dunn D, et al. (2014) The contribution of viral genotype to plasma viral set-point in HIV infection. *PLoS Pathog* 10: e1004112. doi: [10.1371/journal.ppat.1004112](https://doi.org/10.1371/journal.ppat.1004112) PMID: [24789308](https://pubmed.ncbi.nlm.nih.gov/24789308/)
23. Müller V, Fraser C, Herbeck JT (2011) A strong case for viral genetic factors in HIV virulence. *Viruses* 3: 204–16. doi: [10.3390/v3030204](https://doi.org/10.3390/v3030204) PMID: [21994727](https://pubmed.ncbi.nlm.nih.gov/21994727/)
24. Fraser C, Lythgoe K, Leventhal GE, Shirreff G, Hollingsworth TD, et al. (2014) Virulence and pathogenesis of HIV-1 infection: an evolutionary perspective. *Science* 343: 1243727. doi: [10.1126/science.1243727](https://doi.org/10.1126/science.1243727) PMID: [24653038](https://pubmed.ncbi.nlm.nih.gov/24653038/)
25. Shirreff G, Pellis L, Laeyendecker O, Fraser C (2011) Transmission selects for HIV-1 strains of intermediate virulence: a modelling approach. *PLoS Comput Biol* 7: e1002185. doi: [10.1371/journal.pcbi.1002185](https://doi.org/10.1371/journal.pcbi.1002185) PMID: [22022243](https://pubmed.ncbi.nlm.nih.gov/22022243/)
26. Easterling MR, Ellner SP, Dixon PM (2000) Size-specific sensitivity: Applying a new structured population model. *Ecology* 81: 694–708. doi: [10.1890/0012-9658\(2000\)081%5B0694:SSSAAN%5D2.0.CO;2](https://doi.org/10.1890/0012-9658(2000)081%5B0694:SSSAAN%5D2.0.CO;2)

27. Ellner SP, Rees M (2006) Integral Projection Models for Species with Complex Demography. *Am Nat* 167: 410–428. doi: [10.1086/499438](https://doi.org/10.1086/499438) PMID: [16673349](https://pubmed.ncbi.nlm.nih.gov/16673349/)
28. Coulson T (2012) Integral projections models, their construction and use in posing hypotheses in ecology. *Oikos* 121: 1337–1350. doi: [10.1111/j.1600-0706.2012.00035.x](https://doi.org/10.1111/j.1600-0706.2012.00035.x)
29. Carlson JM, Schaefer M, Monaco DC, Batorsky R, Claiborne DT, et al. (2014) HIV transmission. Selection bias at the heterosexual HIV-1 transmission bottleneck. *Science* 345: 1254031. doi: [10.1126/science.1254031](https://doi.org/10.1126/science.1254031) PMID: [25013080](https://pubmed.ncbi.nlm.nih.gov/25013080/)
30. Lythgoe KA, Pellis L, Fraser C (2013) Is HIV short-sighted? Insights from a multistrain nested model. *Evolution* 67: 2769–82. doi: [10.1111/evo.12166](https://doi.org/10.1111/evo.12166) PMID: [24094332](https://pubmed.ncbi.nlm.nih.gov/24094332/)
31. Herbeck JT, Müller V, Maust BS, Ledergerber B, Torti C, et al. (2012) Is the virulence of HIV changing? A meta-analysis of trends in prognostic markers of HIV disease progression and transmission. *AIDS* 26: 193–205. doi: [10.1097/QAD.0b013e32834db418](https://doi.org/10.1097/QAD.0b013e32834db418) PMID: [22089381](https://pubmed.ncbi.nlm.nih.gov/22089381/)
32. Williams BG (2011) Determinants of sexual transmission of HIV: implications for control. arXiv:11084715.
33. Shankarappa R, Margolick JB, Gange SJ, Rodrigo AG, Upchurch D, et al. (1999) Consistent viral evolutionary changes associated with the progression of human immunodeficiency virus type 1 infection. *J Virol* 73: 10489–502. PMID: [10559367](https://pubmed.ncbi.nlm.nih.gov/10559367/)
34. Bartha I, Simon P, Müller V (2008) Has HIV evolved to induce immune pathogenesis? *Trends Immunol* 29: 322–8. doi: [10.1016/j.it.2008.04.005](https://doi.org/10.1016/j.it.2008.04.005) PMID: [18524680](https://pubmed.ncbi.nlm.nih.gov/18524680/)
35. Hool A, Leventhal GE, Bonhoeffer S (2013) Virus-induced target cell activation reconciles set-point viral load heritability and within-host evolution. *Epidemics* 5: 174–80. doi: [10.1016/j.epidem.2013.09.002](https://doi.org/10.1016/j.epidem.2013.09.002) PMID: [24267873](https://pubmed.ncbi.nlm.nih.gov/24267873/)
36. Childs DZ, Rees M, Rose KE, Grubb PJ, Ellner SP (2004) Evolution of size-dependent flowering in a variable environment: construction and analysis of a stochastic integral projection model. *Proc Biol Sci* 271: 425–34. doi: [10.1098/rspb.2003.2597](https://doi.org/10.1098/rspb.2003.2597) PMID: [15101702](https://pubmed.ncbi.nlm.nih.gov/15101702/)
37. Coulson T, MacNulty DR, Stahler DR, vonHoldt B, Wayne RK, et al. (2011) Modeling effects of environmental change on wolf population dynamics, trait evolution, and life history. *Science* 334: 1275–8. doi: [10.1126/science.1209441](https://doi.org/10.1126/science.1209441) PMID: [22144626](https://pubmed.ncbi.nlm.nih.gov/22144626/)
38. Bartha I, Carlson JM, Brumme CJ, McLaren PJ, Brumme ZL, et al. (2013) A genome-to-genome analysis of associations between human genetic variation, HIV-1 sequence diversity, and viral control. *Elife* 2: e01123. doi: [10.7554/eLife.01123](https://doi.org/10.7554/eLife.01123) PMID: [24171102](https://pubmed.ncbi.nlm.nih.gov/24171102/)
39. Falconer DS (1981) Introduction to quantitative genetics. Longman (New York), 2nd edition.