Dissecting HIV Virulence: Heritability of Setpoint Viral Load, CD4+ T Cell Decline and Per-Parasite Pathogenicity

Frederic Bertels,[1] Alex Marzel,[2,3,†] Gabriel Leventhal,[1,†,‡] Venelin Mitov,[4] Jacques Fellay,[5] Huldrych F Günthard,[2,3] Jürg Böni,[3] Sabine Yerly,[6] Thomas Klimkait,[7] Vincent Aubert,[8] Manuel Battegay,[9] Andri Rauch,[10] Matthias Cavassini,[11] Alexandra Calmy,[12] Enos Bernasconi,[13] Patrick Schmid,[14] Alexandra U Scherrer,[2,3] Viktor Müller,[15.16] Sebastian Bonhoeffer, Roger Kouyos,[2,3] Roland R Regoes,[*,1] and the Swiss HIV Cohort Study[17]

[1]Institute of Integrative Biology, ETH Zurich, Zurich, Switzerland
[2]Division of Infectious Diseases and Hospital Epidemiology, University Hospital Zurich, Zurich, Switzerland
[3]Institute of Medical Virology, University of Zurich, Zurich, Switzerland
[4]Department of Biosystems Science and Engineering, ETH Zurich, Zurich, Switzerland
[5]School of Life Sciences, École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland
[6]Laboratory of Virology, Division of Infectious Diseases, Geneva University Hospital, Geneva, Switzerland
[7]Molecular Virology, Department of Biomedicine – Petersplatz, University of Basel, Basel, Switzerland
[8]Division of Immunology and Allergy, University Hospital Lausanne, Lausanne, Switzerland
[9]Division of Infectious Diseases and Hospital Epidemiology, University Hospital Basel, Basel, Switzerland
[10]Department of Infectious Diseases, Berne University Hospital and University of Berne, Berne, Switzerland
[11]Division of Infectious Diseases, University Hospital Lausanne, Lausanne, Switzerland
[12]HIV/AIDS Unit, Infectious Disease Service, Geneva University Hospital, Geneva, Switzerland
[13]Division of Infectious Diseases, Regional Hospital Lugano, Lugano, Switzerland
[14]Division of Infectious Diseases, Cantonal Hospital St Gallen, St Gallen, Switzerland
[15]Institute of Biology, Eötvös Loránd University, Budapest, Hungary
[16]Evolutionary Systems Research Group, MTA Centre for Ecological Research, Tihany, Hungary
[17]Membership list can be found in the Acknowledgments section
[†]These authors contributed equally to this work
[‡]Current address: Department of Civil and Environmental Engineering, Massachusetts Institute of Technology (MIT), Cambridge, MA, 02139, U.S.A.

Corresponding author: roland.regoes@env.ethz.ch

Parts of this study, in particular the evidence for the heritability of the CD4+ T cell decline, have been presented at EPIDEMICS 2015. This paper has been uploaded in agreement with the authors of another study on a similar topic (Blanquart *et al.*, 2017) that is going to be published in PLOS Biology in the next days.

Abstract

Pathogen strains may differ in virulence because they attain different loads in their hosts, or because they induce different disease-causing mechanisms independent of their load. In evolutionary ecology, the latter is referred to as "per-parasite pathogenicity". Using viral load and CD4+ T cell measures from 2014 HIV-1 subtype B infected individuals enrolled in the Swiss HIV Cohort Study, we investigated if virulence — measured by the rate of decline of CD4+ T cells — and per-parasite pathogenicity are heritable from donor to recipient. We estimated heritability by donor-recipient regressions applied to 196 previously identified transmission pairs, and by phylogenetic methods applied to a transmission tree. Applying the donor-recipient regressions to the CD4+ T cell decline and per-parasite pathogenicity did not yield heritability estimates significantly different from zero. With the phylogenetic mixed model approach, however, we find that the heritability of the decline of CD4+ T cells is 25% (95% CI: 9%–40%) assuming neutral evolution of this trait, or 17% (95% CI: 6%–29%) assuming stabilizing selection. The heritability of per-parasite pathogenicity is estimated as 22% (95% CI: 5%–39%) assuming it evolves neutrally, and 17% (95% CI: 4%–29%) for stabilizing selection. Further, we confirm previous studies that established the heritability of the set-point viral load. Interestingly, the pattern of evolution of all these traits differs significantly from neutrality, and is most consistent with stabilizing selection. Our results suggest that viral genetic factors affect virulence in two ways: indirectly through influencing the set-point viral load, and directly by modulating the per-parasite pathogenicity of the virus.

1 Introduction

One of the most common modulators of virulence is the pathogen load. Higher load often leads to more morbidity, or to faster disease progression or death. Similar to virulence, the load that a pathogen strain attains is determined by both, the host and the pathogen. In evolutionary ecology, hosts that limit virulence by reducing pathogen load are called "resistant", and pathogen strains that attain a high load in their hosts are often termed "virulent".

But virulence is not completely determined by the pathogen's load alone. There are pathogen-load-independent components, which are again influenced by the host and the pathogen. A host that suffers less than average from being infected by a pathogen and carrying a specific load is called "tolerant" (Ayres and Schneider, 2012; Boots, 2008; Boots *et al.*, 2009; Little *et al.*, 2010; Medzhitov *et al.*, 2012; Råberg, 2014; Råberg *et al.*, 2009; Read *et al.*, 2008; Schneider and Ayres, 2008; Simms, 2000). A pathogen strain that causes less than average virulence attaining a specific load is said to have a low "per-parasite pathogenicity" (Råberg, 2014; Råberg and Stjernman, 2012). Fig 1A displays these virulence components diagrammatically.

How can pathogen-load-independent components of virulence be determined? To identify these components, "excess virulence" needs to be measured, i. e. by how much virulence differs from what is expected for a specific pathogen load. Statistically speaking, "excess virulence" is the residual virulence after adjusting for differences in pathogen load. This adjustment can be visualized on fitness-versus-pathogen-load plots (Fig 1B). On such a plot, host types with differing levels of disease tolerance are characterized by different tolerance curves that depict the relationship between virulence and pathogen load (see Fig 1B bottom left). The steepness of this curve is inversely related to disease tolerance. Fig 1C shows how per-parasite pathogenicity can be disentangled from host tolerance in general.

HIV infection provides an ideal example to illustrate this decomposition of virulence. First, in HIV infection pathogen load and virulence are positively correlated: it is well-known that higher set-point viral loads are associated with faster progression toward disease and death (Mellors *et al.*, 1996). However, the correlation between the set-point viral load and the decline of the CD4+ T cells — a good proxy of the rate of disease progression — is not very strong: R^2 values were found to be between 0.05—0.08 in American cohorts (Rodriguez *et al.*, 2006), and 0.05 for the population studied here (Regoes *et al.*, 2014) (although the relationship between set-point viral load and survival time has been reported to be higher (Arnaout *et al.*, 1999).) This indicates that there are factors influencing virulence other than the set-point viral load.

Second, in the context of HIV infection, examples for variation in all these virulence components exist. Human genes conferring "host resistance" in the sense of evolutionary ecology (see above) have been identified: Individuals who carry protective HLA-B alleles have lower set-point viral loads, and progress to disease at a slower rate than people without these alleles (Fellay *et al.*, 2007; Goulder and Watkins, 2008). Across primates, species-specific restriction factors can limit the load of Human and Simian Immunodeficiency Viruses (Kirchhoff, 2009; Zheng *et al.*, 2012).

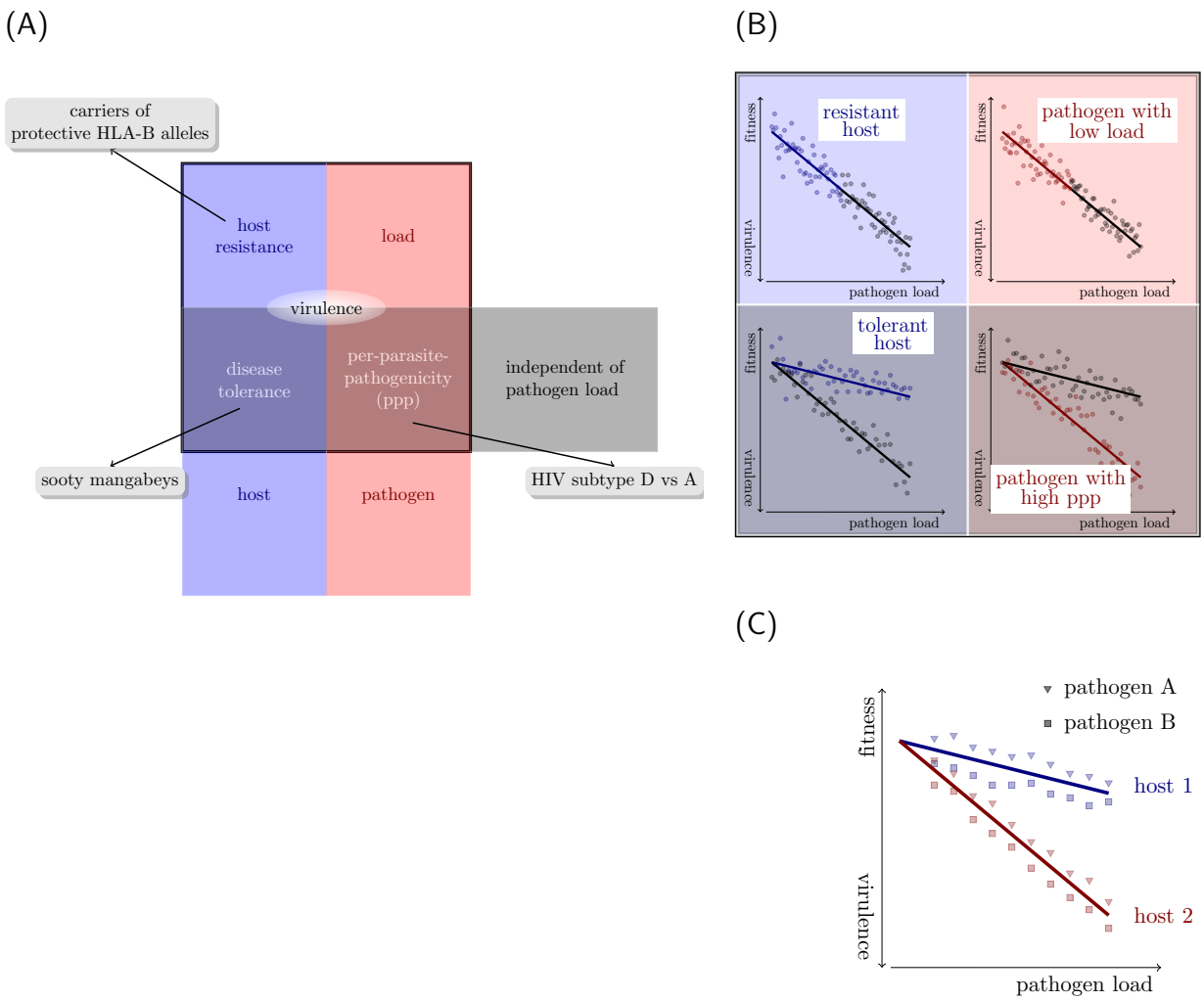

Figure 1: **Dissecting virulence.** (A) Systematics of virulence components. Each component can be a trait of either the pathogen or the host, and depend or be indepependent of the load of the pathogen. (B) Formally, virulence can be dissected using fitness-versus-pathogen-load plots. (In these plots, host fitness is inversely correlated with virulence.) Adapted from Figure 1 in Raberg (Råberg, 2014). (C) In multi-host multi-pathogen systems, virulence components can be disentangled by first defining host-type-specific tolerance curves. Pathogens differing in their per-parasite pathogenicity will then fall on different sides of these tolerance curves. In the example shown, pathogen B has a higher per-parasite pathogenicity than pathogen A.

Third, in the context of immunodeficiency viruses, examples for variation in pathogen-load-independent virulence components exist. Simian Immunodeficiency Virus (SIV) infection in natural hosts, such as the sooty mangabeys, is avirulent despite the high set-point viral loads SIV attains in these hosts (Chahroudi *et al.*, 2012; Chakrabarti, 2004). In contrast, SIV infection in non-natural hosts, such as the rhesus macaque, leads

to an AIDS-like disease (Chahroudi *et al.*, 2012; Chakrabarti, 2004). Thus, natural hosts tolerate the infection without becoming sick.

Lastly, there is also variation in per-parasite pathogenicity across viral strains. For example, there is evidence that HIV-1 subtype D leads to faster disease progression than subtype A even though both subtypes attain similar set-point viral loads (Baeten *et al.*, 2007). In other words, subtype A and D differ in their per-parasite pathogenicity.

Previously, we investigated if humans display variation in disease tolerance against HIV (Regoes *et al.*, 2014). We found that younger individuals and HLA-B heterozygotes are more tolerant, and could link the variation in tolerance to HLA-B genotype. In this previous study, we did however not investigate the potential impact of the virus genotype on the degree of disease tolerance.

How can one investigate if the virus genotype influences a host-pathogen trait? One way is to study associations between genetic polymorphisms of the virus and the trait, as is done, for example, in genome-wide association studies (Bartha *et al.*, 2013). An alternative approach is to estimate the heritability of the trait. If a trait is heritable, i. e. similar between similar viral genotypes, it must, at least in part, be determined by viral genes.

The influence of the virus genotype on the set-point viral load has been the focus of many research groups, including ours, over the past years. Most studies determined the heritability of the set-point viral load (Alizon *et al.*, 2010; Fraser *et al.*, 2014; Hodcroft *et al.*, 2014; Hollingsworth *et al.*, 2010; Müller *et al.*, 2011), while others investigated associations between genetic polymorphisms of the virus and the trait (Bartha *et al.*, 2013). There is a consensus that set-point viral load is heritable, although there is some controversy on the numerical value of the heritability.

In this study, we investigate the influence of the HIV genotype on virulence and its various components by estimating the heritability of these traits. To this end, we use data from the Swiss HIV Cohort Study. For our analysis, we selected cohort participants, for whom we could determine the set-point viral load and the decline of CD4+ T lymphocytes — an established proxy for virulence in HIV infection. As a surrogate for the per-parasite pathogenicity we use the residuals from previously determined tolerance curves (Regoes *et al.*, 2014) as described in Methods and Materials.

Our analysis confirms that set-point viral load is heritable. We further provide evidence that the decline of CD4+ T lymphocytes, i. e. virulence, is heritable. Lastly, we find evidence for the heritability of the per-parasite pathogenicity, the pathogen-load-independent component of virulence. Our results are therefore consistent with the notion that the virus genotype affects virulence in HIV infection both via the viral load, and via viral-load-independent mechanisms.

2 Methods and Materials

2.1 Study population

We analysed a subset of the individuals from the Swiss HIV Cohort Study (www.shcs.ch) (Schoeni-Affolter *et al.*, 2010). This study has enrolled more than 19'000 HIV-infected individuals to date, which constitutes more than 72% of all patients receiving antiretroviral therapy in Switzerland, and is therefore highly representative. The viral load and CD4+ T cell count of each enrolled individual are determined approximately every three months. In some of these individuals, the pol gene of the virus was sequenced.

The study population of the present study consists of a subset of the study population analyzed in a previous study (Regoes *et al.*, 2014). In this previous study, we had included 3036 HIV-1 infected individuals, for whom viral load measurements and CD4+ T cell counts were available to reliably estimate the set-point viral load and CD4+ T cell decline. We restricted our analysis to data obtained before antiretroviral treatment. Furthermore, we excluded the primary and late phases of the infection by discarding measurements during the first 90 days after the estimated date of infection and measurements obtained when the CD4+ T cell count was below $100/\mu L$. Lastly, individuals were included if they had at least two viral load measurements and three CD4+ T cell measurements that were at least 180 days apart.

For the present study, we selected 2014 individuals of the 3036 individuals enrolled previously. Individuals were included if the pol gene of their virus had been sequenced. The genetic information of the virus was necessary for the present study to infer the transmission history and investigate patterns of heritability.

Pol sequence information was obtained from the SHCS genotypic drug resistance database. Sequences are stored in a central database (SmartGene; Integrated Database Network System version 3.6.13). All laboratories perform population-based sequencing (von Wyl *et al.*, 2007; Yang *et al.*, 2015). The drug resistance database includes, in addition to the routinely collected samples, over 11000 samples from the biobank analyzed by systematic retrospective sequencing (Schoeni-Affolter *et al.*, 2010; Yang *et al.*, 2015). The individuals in our study population belong to the following risk groups: MSM – 972 (48%), HET – 435 (21.5%), IDU – 365 (18%), and Other – 252 (12.5%).

The SHCS, enrolling HIV-infected adults aged over 16 years old, has been approved by ethics committees of all participating institutions. The data collection was anonymous and written informed consent was obtained from all participants (Schoeni-Affolter *et al.*, 2010).

2.2 Set-point viral loads and CD4+ T cell declines

For each individual enrolled in our study, the set-point viral load had been determined in a previous study (Regoes *et al.*, 2014) as the geometric mean of the eligible viral load measurements in each individual. Non-detectable viral loads had been set to half the detection limit. The change of CD4+ T cell count over time had previously been estimated as the slope in a linear regression of CD4+ T cell count against the date, at

which they were determined.

Set-point viral loads and CD4+ T cell declines were adjusted for potential covariates by regressing them against sex, age at infection, risk group and ethnicity. Once significant covariates were identified, adjusted traits were defined as the residuals of a regression with these covariates. Subsequent analyses were then conducted with the residuals.

The inclusion criteria, calculation of set-point viral load and CD4+ T cell decline, as well as the model fitting and comparisons had been implemented and performed in the R language of statistical computing (R Core Team, 2013).

2.3 Per-parasite pathogenicity

Per-parasite pathogenicity is defined as the pathogenic potential of a pathogen strain adjusted for its load and host factors that are associated with pathogen-load independent virulence components, i.e. tolerance.

To derive a proxy for the per-parasite pathogenicity of a strain, we first determined the relationship between pathogen load and pathogenicity — called the tolerance curve — for a given host type. In our previous study (Regoes *et al.*, 2014), we found that the age, at which the host was infected, was associated very strongly with the slope of the tolerance curve. Therefore, we determined the tolerance curve specific for the age of the host that harbors the pathogen strain. We did not account for host factors other than age, such as, for example, HLA-B genotype or homozygosity, because this information of host genotype is lacking for the majority of our study population.

In a next step, we predicted the CD4+ T cell decline we should observe given the set-point viral load that this strain attains in the host. Lastly, we calculated by how much the observed CD4+ T cell decline deviates from this prediction (see Fig 2). Formally, this procedure amounts to regressing CD4+ T cell decline against the set-point viral load, adjusting for the age at infection, and calculating the residual of an individual's trait from the regression line.

2.4 Transmission pairs and tree

The transmission pairs were identified as monophyletic clusters on a previous HIV transmission tree (Kouyos *et al.*, 2014). Of these previously established transmission pairs, 196 were present in our study population. The direction of transmission cannot be inferred in these pairs.

The inference of the transmission tree relies on pol gene sequencing of the virus carried by the study subjects. In particular, we had sequences of pol extending over the HXB2 positions 2253–3870, comprising the protease and the reverse transcriptase. All sequences were initially aligned to an HXB2 reference genome (http://www.ncbi.nlm.nih.gov/nuccore/K03455.1) using MUSLCE. We selected the earliest sequence if more than one sequence was available for a person.

To reconstruct the transmission history, we first removed insertions relatively to HXB2. To exclude signatures of parallel evolution due to drug pressure that can distort

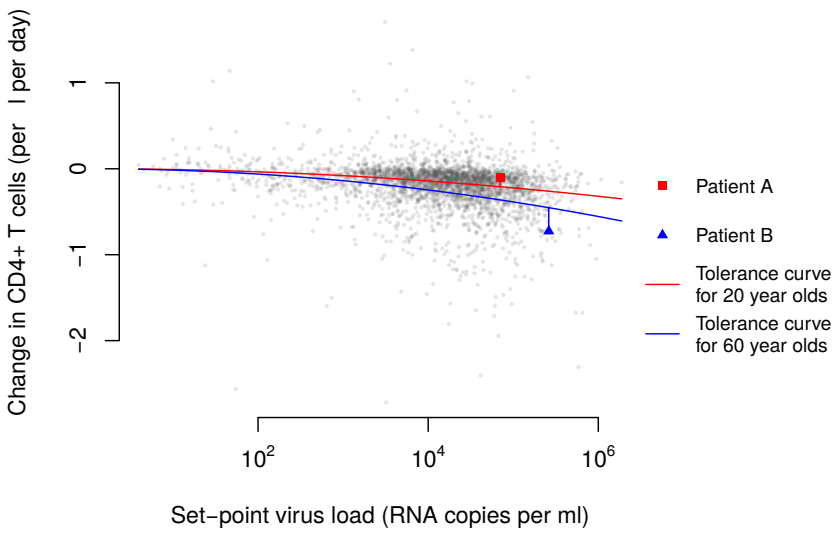

Figure 2: **As a surrogate for the per-parasite pathogenicity we use the residuals from age-adjusted tolerance curves.** In the graph, we plotted each individual's CD4+ T cell decline versus his/her set-point viral load. The red and blue curves show the average relationships between these two measures in the groups that were 20 and 60 years old at the time of their infection, respectively. These age-adjusted tolerance curves were determined previously (Regoes *et al.*, 2014). The red square and blue triangle highlight two individuals with an age at infection of 20 and 60 years, respectively.

the inferred transmission history, we further removed drug resistance mutations according to the databases of Stanford (http://hivdb.stanford.edu/) and the International Antiviral Society (https://www.iasusa.org/). We used Gblocks to refine the alignment. The final number of positions was 1106.

We constructed the transmission tree using FastTree (Version 2.1.8 SSE3, OpenMP). We used a maximum-likelihood-based inference using a Generalized Time-Reversible evolutionary model and a CAT model (Stamatakis, 2006) with 20 discrete evolutionary rate categories. We use the most rigorous and time-consuming FastTree parameters (FastTreeMP -pseudo -spr 4 -mlacc 2 -slownni -gtr -nt). Alignment bootstrapping was done with SEQBOOT. Node support values were generated based on 100 bootstrapped trees and the reference (non-bootstrapped alignment) tree using the script "CompareToBootstrap". We rooted the tree with 10 Subtype C sequences as an outgroup, using the R package APE. The tree (without the outgroup sequences) is shown in Fig 3. The branch lengths in our tree correspond to genetic distances between the sequences, and not to time.

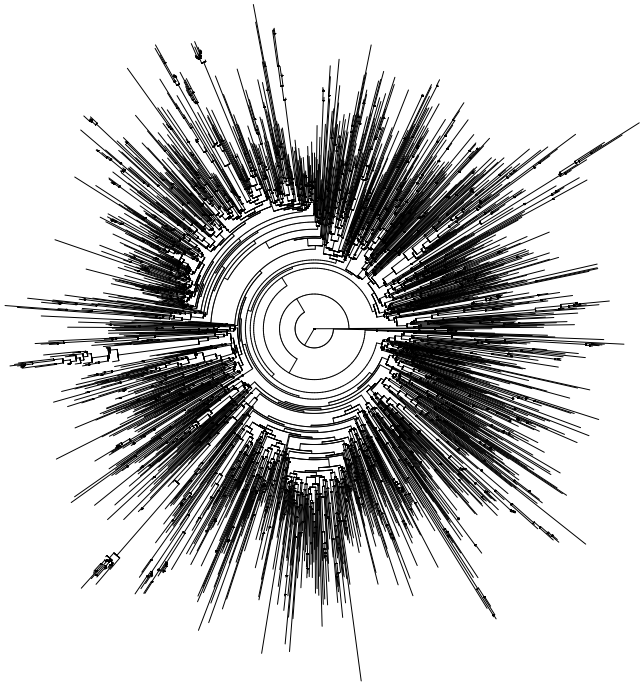

Figure 3: **Rooted phylogenetic tree used in our analysis.** The tree was constructed from HIV-1 pol sequences from 2024 patients of the Swiss HIV Cohort Study. 2014 HIV-1 subtype B sequences were derived from individuals in our study population. Another 10 subtype C sequences were used as a outgroup, but are not shown on the tree.

2.5 Heritability estimation

To estimate the heritability of the three traits — set-point viral load, CD4+ T cell decline and per-parasite pathogenicity — we used two approaches.

First, we applied donor-recipient regressions that are formally equivalent to parent-offspring regressions (Fraser *et al.*, 2014) to the 196 previously identified transmission pairs. Although we do not have any information on the direction of transmission in these pairs, we expect that a regression between the trait in question will yield a good estimate of the heritability. An assessment of how well heritability can be estimated from our undirected transmission pairs can be found in Figure S2.

Second, we employed phylogenetic mixed models (Housworth *et al.*, 2004) that are widely used to estimate the heritability from a phylogenetic tree. These methods have the advantage of being able to incorporate larger study populations and transmission relationships ranging from close pairing to distant epidemiological linkage. We used the recent implementation by Leventhal and Bonhoeffer (Leventhal and Bonhoeffer, 2016) in

the R language of statistical computing (R Core Team, 2013). The models that underlie this method assume trait evolution according to Brownian motion, i.e. that traits drift neutrally. Because we use the method on a tree, the branch lengths of which correspond to genetic distances, we make the implicit assumption that heritability increases linearly with genetic distance. For methodological reasons (Leventhal and Bonhoeffer, 2016), we refrained from using Pagel's λ in our main analysis. However, as a point of comparison with estimates of Pagel's λ in previous studies, we report our estimates of this quantity in the supplementary material. We also applied phylogenetic mixed models based on the Ornstein-Uhlenbeck process that describe stabilizing trait selection around an optimal trait value rather than neutral drift. We used the implementation by Mitov and Stadler Mitov and Stadler (2016).

3 Results

3.1 Heritability of set-point viral load confirmed

The heritability of the set-point viral load has previously been estimated from data of the Swiss HIV cohort study (Alizon *et al.*, 2010) and from data of other cohorts (Hodcroft *et al.*, 2014; Hollingsworth *et al.*, 2010; Müller *et al.*, 2011). The methods differed across these studies.

To test the conclusion of these studies, we applied donor-recipient regressions and the phylogenetic mixed models to the set-point viral loads from the 2014 individuals we included in the present study. The donor-recipient regressions were applied to 196 previously determined transmission pairs (Kouyos *et al.*, 2014), while the phylogenetic methods were applied to a phylogenetic tree that is shown in Fig 3 and was constructed from pol gene sequences (see Methods and Materials). Since set-point viral loads were significantly with sex, age at infection, and were higher in men who have sex with men, we also estimated the heritability of adjusted set-point viral loads as defined in the Methods and Materials.

Across all the methods we use, the estimates for heritability range from 8% to 29% (see Table 1, Fig 4B and Figure S3B. These estimates are all significantly larger than zero, except for the adjusted set-point viral load using the phylogenetic mixed-model that assumes neutral evolution. This result adds to the growing consensus that set-point viral load is heritable.

Interestingly, assuming stabilizing selection on the set-point viral load in our phylogenetic analysis led to a significantly better fit to the data than assuming neutral drift (Likelihood ratio test: $p = 1.2 \times 10^{-4}$ for unadjusted and $p = 8.8 \times 10^{-6}$ for adjusted set-point viral loads). Thus, the estimates for the heritability of the set-point viral load with the best statistical support are 26% and 29% without and with adjustment for cofactors, respectively. Both of these estimates are significantly different from zero.

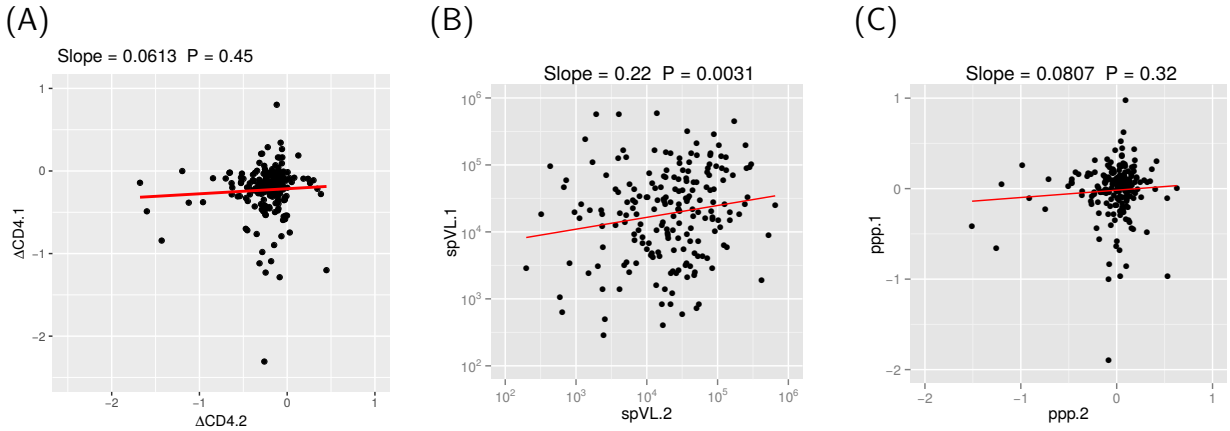

Figure 4: **Heritability estimates from donor-recipient regressions.** For these regressions we plotted the trait values for each partner in the transmission pairs onto the x- and y-axis. Since we do not know the direction of the transmission in the pairs, the assignment of the partners to either x- or y-axis is random. (See also Figure S2 and S3.)

3.2 Evidence for the heritability of CD4+ T cell decline

The set-point viral load is an important determinant of CD4+ T cell decline, and it is heritable. We therefore expect the CD4+ T cell decline to be also heritable "by association". So far, however, there has been no evidence for the heritability of the CD4+ T cell decline or HIV virulence.

To assess the heritability of the CD4+ T cell decline, we applied the same methods as for the set-point viral load. Because the CD4+ T cell decline was associated significantly only with the age at infection we also conducted the analyses with age-adjusted CD4+ T cell declines as defined in the Methods and Materials.

The donor-recipient regression (Fig 4A) results in a hertability estimate, which is not significantly different from zero for both unadjusted and age-adjusted CD4+ T cell declines. This is likely the result of the low nuber of individuals in the transmission pairs (2×196).

Using the phylogenetic mixed models, however, we can incorporate all 2014 individuals of our study population, and obtain a heritability estimates significantly larger than zero. Assuming neutral trait evolution (PMM) yields 25% and 24% for unadjusted and adjusted CD4+ T cell declines, respectively. With stabilizing trait evolution we get 17% irrespective of any adjustment (see Table 1). Again, the assumption of stabilizing selection has more statistical support than neutral drift (Likelihood ratio test: $p = 6.9 \times 10^{-8}$ for unadjusted and $p = 2.6 \times 10^{-5}$ for adjusted CD4+ T cell declines).

Table 1: Heritability estimates for set-point viral load (spVL), CD4+ T cell decline (ΔCD4) and per-parasite pathogenicity (ppp) based on the phylogenetic mixed models assuming Brownian motion-type trait evolution (PMM) or stabilizing trait selection according to the Ornstein-Uhlenbeck process (POUMM). 95% confidence intervals are given in brackets.

	PMM	POUMM
ΔCD4 (unadjusted)	25% (9%–40%)	17% (6%–29%)
ΔCD4 (adjusted)	24% (7%–39%)	17% (5%–30%)
spVL (unadjusted)	12% (2%–28%)	26% (8%–43%)
spVL (adjusted)	8% (0%–26%)	29% (12%–46%)
ppp	22% (5%–39%)	17% (4%–29%)

3.3 Evidence for the heritability of the per-parasite pathogenicity

Lastly, we investigated if there is any evidence for the heritability for the per-parasite pathogenicity. The per-parasite pathogenicity is the component of virulence, which is determined by the pathogen genotype and independent of the pathogen load. Formally, we determine per-parasite pathogenicity by calculating the residual of the regression of the CD4+ T cell decline against the set-point viral load adjusted for the age of the infected individual (see Material and Methods and Fig 2).

Donor-recipient regression (Fig 4C) yields an estimate, which is not significantly different from zero, again likely due to the low number of transmission pairs. Using phylogenetic mixed models, we estimate a statistically significant heritability of 22% assuming neutral drift, and 17% for stabilizing selection (see Table 1). As for the set-point viral load and the CD4+ T cell decline, assuming stabilizing selection led to a significantly better fit than assuming neutral drift (Likelihood ratio test: $p = 2.2 \times 10^{-6}$).

4 Discussion

In this study, we confirmed the heritability of the set-point viral load. Further, we report one of the first pieces of evidence for the heritability of the CD4+ T cell decline, a surrogate of virulence. We also found support for the hypothesis that the pathogen-load-independent virulence component is heritable. Lastly, the evolution of these three traits is significantly better described by the Ornstein-Uhlenbeck process than by Brownian motion.

Our study confirms previous studies that established the heritability of the set-point viral load in HIV infection (as reviewed by Müller *et al.* (2011) and Fraser *et al.* (2014)). In particular, our estimates are consistent with those from a donor-recipient regression in Hollingsworth *et al.* (2010), and a recent analysis by Mitov and Stadler (2016), in which the authors applied their implementation of phylogenetic mixed models based on the

Ornstein-Uhlenbeck process to set-point viral load data. Our analysis is also consistent with the study by Hodcroft *et al.* (2014) that reported a low heritability of 5.7% of the set-point viral load adjusted for covariates and assuming Brownian trait evolution. If we adjust for covariates and assume Brownian trait evolution, we obtain a heritability estimate of 8% that is not significantly different from zero. Assuming trait evolution according to the Ornstein-Uhlenbeck process, however, provides a significantly better fit to the adjusted set-point viral load data and yields a heritability estimate of 29%.

Recently, Carlson et al (Carlson *et al.*, 2016) hypothesized that the differences in the estimates in heritability might be due to the diversity of the Human Leukocyte Antigen (HLA) genes in the various study populations. They showed that restricting the analysis to transmission couples with similar HLA adaptation pressure increases the estimate of the heritability of the set-point viral load. Thus, in a population with diverse HLA alleles heritability will be underestimated and may not even be detectable. Conversely, large heritability estimates may arise in homogeneous study populations. However, the discrepancy between the estimate of Alizon et al and that reported here — both obtained from data of the Swiss HIV Cohort Study with substantial overlap between the two study populations — suggest that methodological aspects, rather than the composition of the study populations, explain the wide range of the heritability estimates.

We find clear evidence for the heritability of the CD4+ T cell decline. As the level of CD4+ T cells is a defining characteristic of clinical AIDS, the CD4+ T cell decline is a good surrogate of virulence of HIV infection. Although the potential heritability of the rate of decline of CD4+ T cells has been investigated previously (Alizon *et al.*, 2010), it was found to be not significantly different from zero. We attribute this discrepancy to the low sample size of the earlier study. In contrast to the 2014 individuals in our study population, Alizon et al had enrolled only 1100 and investigated the heritability only in subpopulations consisting of a few hundered individuals.

We also provide evidence for the heritability of the per-parasite pathogenicity. This trait describes the pathogenic potential of a viral strain that is independent of the load the strain attains in its host. We approximated this trait as the deviation of the CD4+ T cell decline observed in an individual from that predicted on the basis of the observed set-point viral load and the age of the infected host (see Methods). In addition to being determined by per-parasite pathogenicity, this deviation could be affected by further host factors, other sources of biological variation, and, of course, measurement noise, and should therefore be considered only as a surrogate for the per-parasite pathogenicity of a strain. It is important to note, however, that the uncertainties surrounding the quantification of the per-parasite pathogenicity make it more difficult to establish the heritability of this trait. The fact that we found evidence for heritability means that the signal in our surrogate measure of the per-parasite pathogenicity is not completely clouded by factors, for which we could not account.

The heritability of this trait means that there are viral genes that influence the CD4+ T cell decline in ways that does not depend on the viral load. One conceivable such mechanism could be that viral genotypes with high per-parasite pathogenicity elicit ineffective immune responses that, rather than reducing viral load, accelerate CD4+ T cell decline. Because our heritability analysis cannot identify which viral genes are

involved in this load-independent pathogenicity and cannot pinpoint specific mechanisms further research into these aspect is needed.

Previously, Pagel's λ was also employed to estimate the heritability of the set-point viral load from a phylogenetic tree of HIV pol sequences (Alizon *et al.*, 2010). Leventhal and Bonhoeffer (Leventhal and Bonhoeffer, 2016), however, have argued recently that Pagel's λ implicitly assumes the trees to be ultrametric. Thus, for non-ultrametric transmission trees, such as the one we analyzed, this essential assumption of Pagel's λ is violated, and this method should therefore not be used. We nevertheless provide heritability estimates for comparison in Figure S1. With Pagel's λ the set-point viral load and the CD4+ T cell decline are also found to be significantly heritable.

In summary, we presented a comprehensive evolutionary analysis of HIV virulence. We established that viral load dependent and independent virulence components, as well as overall virulence are heritable. This strongly suggests that these virulence components are under partial genetic control of the virus. Future research will need to identify the specific genetic polymorphisms associated with these virulence components.

5 Acknowledgments

We thank the patients who participate in the Swiss HIV Cohort Study (SHCS); the physicians and study nurses, for excellent patient care; the resistance laboratories, for high-quality genotypic drug resistance testing; SmartGene (Zug, Switzerland), for technical support; Brigitte Remy, RN, Martin Rickenbach, MD, Franziska Schöni-Affolter, MD, and Yannick Vallet, MSc, from the SHCS Data Center (Lausanne, Switzerland), for data management; and Danièle Perraudin and Mirjam Minichiello for administrative assistance.

RRR acknowledges the financial support of the Swiss National Science Foundation (grant number: 31003A_149769). This study has been performed within the framework of the Swiss HIV Cohort Study, supported by the Swiss National Science Foundation (grant number 33CS30_148522) and was further supported by the SHCS research foundation. The Swiss HIV Drug Resistance database is supported by SNF project 320030_159868 to H. F. G, the Yvonne-Jacob Foundation; Gilead, Switzerland (1 unrestricted grant to the SHCS Research Foundation and 1 unrestricted grant to H. F. G.); and the University of Zurich's Clinical Research Priority Program (Viral Infectious Diseases: Zurich Primary HIV Infection Study; to H. F. G.)

The data are gathered by the 5 Swiss University Hospitals, 2 Cantonal Hospitals, 15 affiliated hospitals and 36 private physicians (listed in http://www.shcs.ch/31-health-care-providers). The members of the Swiss HIV Cohort Study are: Aubert V, Barth J, Battegay M, Bernasconi E, Böni J, Bucher HC, Burton-Jeangros C, Calmy A, Cavassini M, Egger M, Elzi L, Fehr J, Fellay J, Furrer H (Chairman of the Clinical and Laboratory Committee), Fux CA, Gorgievski M, Günthard H (President of the SHCS), Haerry D (deputy of "Positive Council"), Hasse B, Hirsch HH, Hösli I, Kahlert C, Kaiser L, Keiser O, Klimkait T, Kouyos R, Kovari H, Ledergerber B, Martinetti G, Martinez de Tejada B, Metzner K, Müller N, Nadal D, Pantaleo G, Rauch A (Chairman of the Scientific

Board), Regenass S, Rickenbach M (Head of Data Center), Rudin C (Chairman of the Mother & Child Substudy), Schöni-Affolter F, Schmid P, Schultze D, Schüpbach J, Speck R, Staehelin C, Tarr P, Telenti A, Trkola A, Vernazza P, Weber R, Yerly S.

References

Alizon, S., von Wyl, V., Stadler, T., Kouyos, R. D., Yerly, S., Hirschel, B., Böni, J., Shah, C., Klimkait, T., Furrer, H., Rauch, A., Vernazza, P. L., Bernasconi, E., Battegay, M., Bürgisser, P., Telenti, A., Günthard, H. F., and Bonhoeffer, S. 2010. Phylogenetic approach reveals that virus genotype largely determines HIV set-point viral load. *PLoS Pathog.*, 6(9): e1001123.

Arnaout, R. A., Lloyd, A. L., O'Brien, T. R., Goedert, J. J., Leonard, J. M., and Nowak, M. A. 1999. A simple relationship between viral load and survival time in HIV-1 infection. *Proc. Natl. Acad. Sci. USA*, 96(20): 11549–11553.

Ayres, J. S. and Schneider, D. S. 2012. Tolerance of infections. *Annu. Rev. Immunol.*, 30: 271–94.

Baeten, J. M., Chohan, B., Lavreys, L., Chohan, V., McClelland, R. S., Certain, L., Mandaliya, K., Jaoko, W., and Overbaugh, J. 2007. HIV-1 subtype D infection is associated with faster disease progression than subtype A in spite of similar plasma HIV-1 loads. *J. Infect. Dis.*, 195(8): 1177–80.

Bartha, I., Carlson, J. M., Brumme, C. J., McLaren, P. J., Brumme, Z. L., John, M., Haas, D. W., Martinez-Picado, J., Dalmau, J., López-Galíndez, C., Casado, C., Rauch, A., Günthard, H. F., Bernasconi, E., Vernazza, P., Klimkait, T., Yerly, S., O'Brien, S. J., Listgarten, J., Pfeifer, N., Lippert, C., Fusi, N., Kutalik, Z., Allen, T. M., Müller, V., Harrigan, P. R., Heckerman, D., Telenti, A., and Fellay, J. 2013. A genome-to-genome analysis of associations between human genetic variation, HIV-1 sequence diversity, and viral control. *Elife*, 2: e01123.

Blanquart, F., Wymant, C., Cornelissen, M., Gall, A., Bakker, M., Bezemer, D., Hall, M., Hillebregt, M., Ong, S. H., Albert, J., Bannert, N., Fellay, J., Fransen, K., Gourlay, A., Grabowski, M. K., Gunsenheimer-Bartmeyer, B., Günthard, H. F., Kivelä, P., Kouyos, R., Laeyendecker, O., Liitsola, K., Meyer, L., Porter, K., Ristola, M., van Sighem, A., Vanham, G., Berkhout, B., Kellam, P., Reiss, P., and Fraser, C. on behalf of the BEEHIVE collaboration 2017. Viral genetic variation accounts for a third of variability in HIV-1 set-point viral load in Europe. *PLoS Biol.*, in press.

Boots, M. 2008. Fight or learn to live with the consequences? *Trends Ecol. Evol. (Amst.)*, 23(5): 248–50.

Boots, M., Best, A., Miller, M. R., and White, A. 2009. The role of ecological feedbacks in the evolution of host defence: what does theory tell us? *Philos. Trans. R. Soc. Lond., B, Biol. Sci.*, 364(1513): 27–36.

Carlson, J. M., Du, V. Y., Pfeifer, N., Bansal, A., Tan, V. Y. F., Power, K., Brumme, C. J., Kreimer, A., DeZiel, C. E., Fusi, N., Schaefer, M., Brockman, M. A., Gilmour, J., Price, M. A., Kilembe, W., Haubrich, R., John, M., Mallal, S., Shapiro, R., Frater, J., Harrigan, P. R., Ndung'u, T., Allen, S., Heckerman, D., Sidney, J., Allen, T. M., Goulder, P. J. R., Brumme, Z. L., Hunter, E., and Goepfert, P. A. 2016. Impact of pre-adapted HIV transmission. *Nat. Med.*, 22(6): 606–13.

Chahroudi, A., Bosinger, S. E., Vanderford, T. H., Paiardini, M., and Silvestri, G. 2012. Natural SIV hosts: showing AIDS the door. *Science*, 335(6073): 1188–93.

Chakrabarti, L. A. 2004. The paradox of simian immunodeficiency virus infection in sooty mangabeys: active viral replication without disease progression. *Front Biosci*, 9: 521–39.

Fellay, J., Shianna, K. V., Ge, D., Colombo, S., Ledergerber, B., Weale, M., Zhang, K., Gumbs, C., Castagna, A., Cossarizza, A., Cozzi-Lepri, A., De Luca, A., Easterbrook, P., Francioli, P., Mallal, S., Martinez-Picado, J., Miro, J. M., Obel, N., Smith, J. P., Wyniger, J., Descombes, P., Antonarakis, S. E., Letvin, N. L., McMichael, A. J., Haynes, B. F., Telenti, A., and Goldstein, D. B. 2007. A whole-genome association study of major determinants for host control of HIV-1. *Science*, 317(5840): 944–7.

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

Goulder, P. J. R. and Watkins, D. I. 2008. Impact of MHC class I diversity on immune control of immunodeficiency virus replication. *Nat. Rev. Immunol.*, 8(8): 619–30.

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

Hollingsworth, T. D., Laeyendecker, O., Shirreff, G., Donnelly, C. A., Serwadda, D., Wawer, M. J., Kiwanuka, N., Nalugoda, F., Collinson-Streng, A., Ssempijja, V., Hanage, W. P., Quinn, T. C., Gray, R. H., and Fraser, C. 2010. HIV-1 transmitting couples have similar viral load set-points in Rakai, Uganda. *PLoS Pathog.*, 6(5): e1000876.

Housworth, E. A., Martins, E. P., and Lynch, M. 2004. The phylogenetic mixed model. *Am. Nat.*, 163(1): 84–96.

Kirchhoff, F. 2009. Is the high virulence of HIV-1 an unfortunate coincidence of primate lentiviral evolution? *Nat. Rev. Microbiol.*, 7(6): 467–76.

Kouyos, R. D., Rauch, A., Böni, J., Yerly, S., Shah, C., Aubert, V., Klimkait, T., Kovari, H., Calmy, A., Cavassini, M., Battegay, M., Vernazza, P. L., Bernasconi, E., Ledergerber, B., and Günthard, H. F. 2014. Clustering of HCV coinfections on HIV phylogeny indicates domestic and sexual transmission of HCV. *Int J Epidemiol*, 43(3): 887–96.

Leventhal, G. E. and Bonhoeffer, S. 2016. Potential Pitfalls in Estimating Viral Load Heritability. *Trends Microbiol.*

Little, T. J., Shuker, D. M., Colegrave, N., Day, T., and Graham, A. L. 2010. The coevolution of virulence: tolerance in perspective. *PLoS Pathog.*, 6(9): e1001006.

Medzhitov, R., Schneider, D. S., and Soares, M. P. 2012. Disease tolerance as a defense strategy. *Science*, 335(6071): 936–41.

Mellors, J. W., Rinaldo, C. R., Gupta, P., White, R. M., Todd, J. A., and Kingsley, L. A. 1996. Prognosis in HIV-1 infection predicted by the quantity of virus in plasma. *Science*, 272(5265): 1167–1170.

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

Müller, V., Fraser, C., and Herbeck, J. T. 2011. A strong case for viral genetic factors in HIV virulence. *Viruses*, 3(3): 204–16.

R Core Team 2013. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Råberg, L. 2014. How to live with the enemy: understanding tolerance to parasites. *PLoS Biol.*, 12(11): e1001989.

Råberg, L. and Stjernman, M. 2012. *Ecoimmunology*, chapter The evolutionary ecology of infectious disease virulence., page 548578. Oxford University Press.

Råberg, L., Graham, A. L., and Read, A. F. 2009. Decomposing health: tolerance and resistance to parasites in animals. *Philos. Trans. R. Soc. Lond., B, Biol. Sci.*, 364(1513): 37–49.

Read, A. F., Graham, A. L., and Råberg, L. 2008. Animal defenses against infectious agents: is damage control more important than pathogen control. *PLoS Biol.*, 6(12): e4.

Regoes, R. R., McLaren, P. J., Battegay, M., Bernasconi, E., Calmy, A., Günthard, H. F., Hoffmann, M., Rauch, A., Telenti, A., and Fellay, J. 2014. Disentangling Human Tolerance and Resistance Against HIV. *PLoS Biol.*, 12(9): e1001951.

Rodriguez, B., Sethi, A. K., Cheruvu, V. K., Mackay, W., Bosch, R. J., Kitahata, M., Boswell, S. L., Mathews, W. C., Bangsberg, D. R., Martin, J., Whalen, C. C., Sieg, S., Yadavalli, S., Deeks, S. G., and Lederman, M. M. 2006. Predictive value of plasma HIV RNA level on rate of CD4 T-cell decline in untreated HIV infection. *JAMA*, 296(12): 1498–506.

Schneider, D. S. and Ayres, J. S. 2008. Two ways to survive infection: what resistance and tolerance can teach us about treating infectious diseases. *Nat. Rev. Immunol.*, 8(11): 889–95.

Schoeni-Affolter, F., Ledergerber, B., Rickenbach, M., Rudin, C., Günthard, H. F., Telenti, A., Furrer, H., Yerly, S., and Francioli, P. 2010. Cohort profile: the Swiss HIV Cohort study. *Int J Epidemiol*, 39(5): 1179–89.

Simms, E. L. 2000. Defining tolerance as a norm of reaction. *Evolutionary Ecology*, 14(4-6): 563–570.

Stamatakis, A. 2006. Phylogenetic models of rate heterogeneity: a high performance computing perspective. In *Proceedings 20th IEEE International Parallel & Distributed Processing Symposium*, pages 8–pp. IEEE.

von Wyl, V., Yerly, S., Böni, J., Bürgisser, P., Klimkait, T., Battegay, M., Furrer, H., Telenti, A., Hirschel, B., Vernazza, P. L., Bernasconi, E., Rickenbach, M., Perrin, L., Ledergerber, B., and Günthard, H. F. 2007. Emergence of HIV-1 drug resistance in previously untreated patients initiating combination antiretroviral treatment: a comparison of different regimen types. *Arch. Intern. Med.*, 167(16): 1782–90.

Yang, W.-L., Kouyos, R., Scherrer, A. U., Böni, J., Shah, C., Yerly, S., Klimkait, T., Aubert, V., Furrer, H., Battegay, M., Cavassini, M., Bernasconi, E., Vernazza, P., Held, L., Ledergerber, B., and Günthard, H. F. 2015. Assessing the Paradox Between Transmitted and Acquired HIV Type 1 Drug Resistance Mutations in the Swiss HIV Cohort Study From 1998 to 2012. *J. Infect. Dis.*, 212(1): 28–38.

Zheng, Y.-H., Jeang, K.-T., and Tokunaga, K. 2012. Host restriction factors in retroviral infection: promises in virus-host interaction. *Retrovirology*, 9: 112.

Supplementary Figures

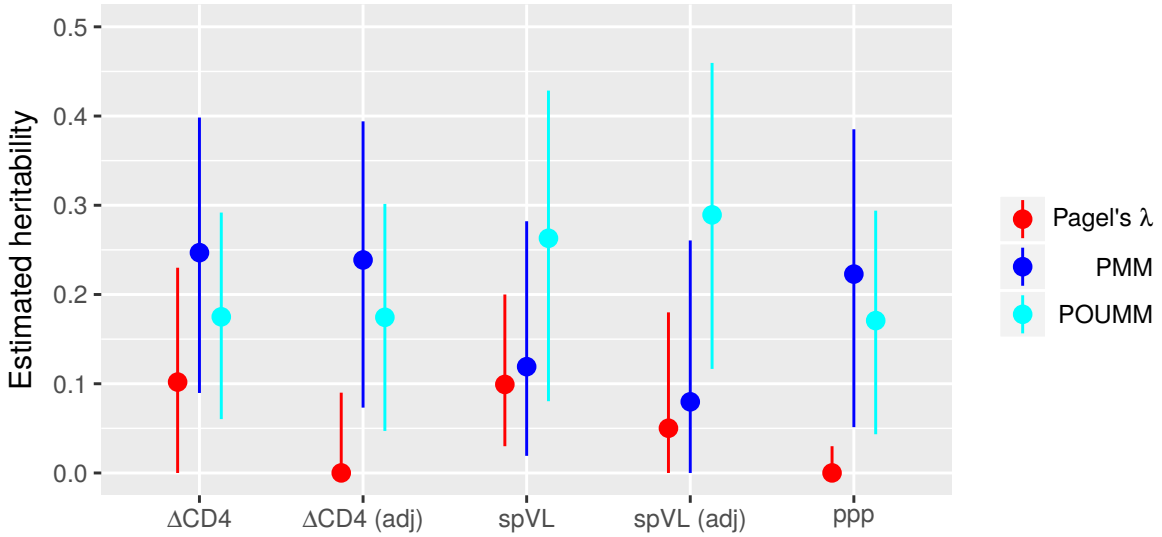

Figure S1: Heritability estimates using using Pagel's λ in comparison to those using PMM and POUMM. The points are the point estimates, the vertical lines show the 95% confidence intervals. We show these estimates only for comparison with older papers that used Pagel's λ on unadjusted traits. Because of the non-ultrametric nature of our phylogenetic tree Pagel's λ is not appropriate to use.

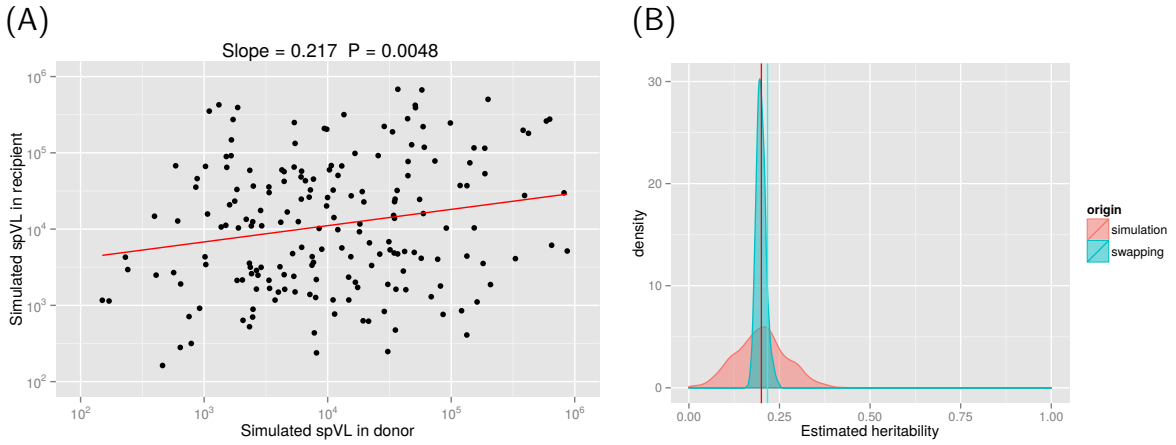

Figure S2: Randomly swapping donors and recipients does not affect the heritability estimate significantly. For this analysis we assumed a true heritability of 0.20 and 200 donor-recipient pairs. (A) Output of a single simulation. The estimated heritability is 0.217 and significantly different from zero. (B) Simulating 1000 datasets with 200 donor-recipient pairs and determining the slope of the regression results in heritability estimates shown by the density in red with a mean shown by the red vertical line. Randomly swapping donor and recipient in one of the simulated datasets leads to estimates shown by the cyan density, which is much narrower than the expected variance of the donor-recipient regression slope. Thus, ignoring the direction of transmission in the transmission pairs we analyzed introduces far less uncertainty into the heritability estimate than we get by randomly selecting 200 pairs.

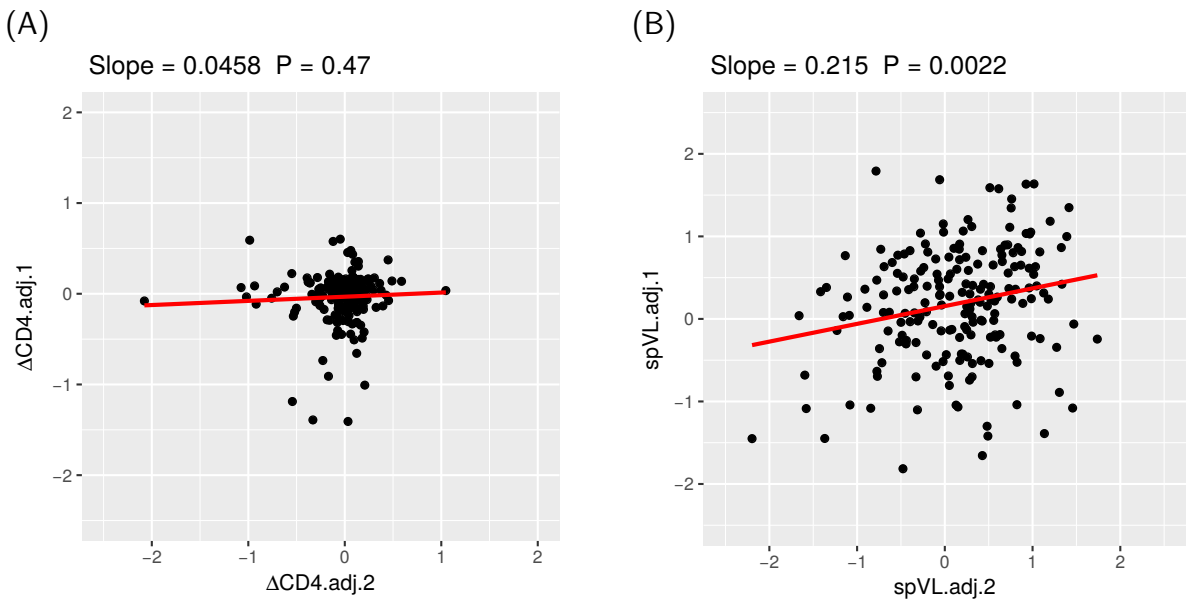

Figure S3: Heritability estimates from donor-recipient regressions on adjusted CD4+ T cell decline and set-point viral loads. For the definition of the adjusted trait values see the Methods and Materials. (See also Fig 4.)