CellPress

Opinion

Potential Pitfalls in Estimating Viral Load Heritability

Gabriel E. Leventhal[1,2,*] and Sebastian Bonhoeffer[2,*]

In HIV patients, the set-point viral load (SPVL) is the most widely used predictor of disease severity. Yet SPVL varies over several orders of magnitude between patients. The heritability of SPVL quantifies how much of the variation in SPVL is due to transmissible viral genetics. There is currently no clear consensus on the value of SPVL heritability, as multiple studies have reported apparently discrepant estimates. Here we illustrate that the discrepancies in estimates are most likely due to differences in the estimation methods, rather than the study populations. Importantly, phylogenetic estimates run the risk of being strongly confounded by unrealistic model assumptions. Care must be taken when interpreting and comparing the different estimates to each other.

Searching for SPVL Heritability in HIV Infection

During the **asymptomatic phase** (see Glossary) of HIV infection, the viral load within a patient fluctuates around a relatively stable value known as the set-point viral load (SPVL). SPVL has proven to be highly relevant, as untreated patients with higher SPVL tend to progress to AIDS faster than those with low SPVL, and consequentially SPVL is one of the most widely used predictors of disease severity [1,2]. What is particularly striking, however, is the amount of variation in SPVL between patients: SPVL varies over several orders of magnitude between patients [1,3–5]. Understanding the source of this wide variation in SPVL in the patient population is key to understanding HIV pathogenicity and why certain patients progress to AIDS rather quickly, while others progress much more slowly or not at all.

The potential contributions to SPVL variation can generally be split into four categories, determined by:

(i) the specific genetic viral variant that infects a host;
(ii) the host genetics;
(iii) any interactions between host and viral genetics;
(iv) other extrinsic factors, both independent of, and interacting with, host and viral genetics.

From a virus-centric view, we can group contributions (ii)–(iv) together, and hence the question becomes to what degree the variation in HIV genetics explains the variation in SPVL in the patient population, and by proxy, how much viral genetics controls the variation in HIV pathogenicity.

New HIV infections can occur after a virus is transmitted from a donor host to a recipient host. Thus, from an evolutionary point of view, viral genetic information is passed on and conserved from one infection to the next, whereas donor and recipient host genetics are typically unrelated. In evolutionary theory, the concept of heritability quantifies precisely the question at hand [6]: how much of the observed variation in a trait is explained by variation in the genetics that are passed on to the next generation? A heritability of 100% means that all of the trait variation is explained by transmissible genetic information, while a heritability of 0% means that

Trends

Set-point viral load (SPVL) is the most used predictor of HIV disease severity.

There is an ongoing debate about the importance of host and viral genetic factors to SPVL.

The increasing availability of HIV sequence data allows for the use of potentially powerful new phylogenetic tools to investigate the genetic underpinnings of SPVL.

The heritability of SPVL in a population can quantify the relative influence of virus and host on the variation in SPVL in the population, but the current reported estimates using a variety of methods, among others phylogenetic methods, are seemingly discordant.

In order to make sense of the discrepant estimates, it is necessary to disentangle biases due to the different methods used from true differences in the study populations.

[1]Department of Civil and Environmental Engineering, Massachusetts Institute of Technology (MIT), Cambridge, MA 02143, USA
[2]Institute of Integrative Biology, ETH Zurich, 8092 Zurich, Switzerland

*Correspondence: gaberoo@mit.edu (G.E. Leventhal) and sebastian.bonhoeffer@env.ethz.ch (S. Bonhoeffer).

transmissible genetics explains none of the trait variation. A number of recent publications have reported heritability estimates based on different methods that range from as low as 5% to as high as 50% [7–14]. These studies used partly different methods to estimate heritability, depending on the available data. The crucial question thus arises as to whether these discrepant estimates reflect true differences in the study populations, or whether the differences may rather be artifacts of the estimation method used.

Every statistical model comes with a set of assumptions that must be fulfilled in order for the results to be meaningful. Here, we aim to illustrate how some of these estimates of heritability might be confounded by model assumptions that do not completely conform to the dynamics of HIV transmission. Nevertheless, even when the assumptions are fulfilled, the choice of estimation method alone can potentially lead to differences in the heritability estimates.

Estimating the Heritability of a Viral Trait

The concept of heritability was initially developed to measure the degree of correspondence between the trait value in an offspring and the trait value of its parents, and can be traced back to Galton [15], though the first use of the term 'heritability' remains elusive [16]. Most of the credit for developing the methods to estimate heritability go back to Fisher [17] and Wright [18], although the common contemporary use corresponds to 'narrow sense heritability' as defined by Lush [19]: heritability is the ratio of additive genetic variance, σ_g^2, and the sum of additive genetic and environmental variance, σ_e^2, in a population,

$$h^2 = \frac{\sigma_g^2}{\sigma_g^2 + \sigma_e^2},$$

where the total phenotypic variance, $\sigma_\phi^2 = \sigma_g^2 + \sigma_e^2$, is the sum of all contributions to variance. Heritability can then be estimated from data of related individuals (e.g., parents and their offspring) using regression methods [6,17,18]. For the case of SPVL in HIV, σ_g^2 corresponds to the transmissible contribution (i) and σ_e^2 comprises the nontransmissible contributions (ii)–(iv).

Analogous to these methods that were developed for sexually reproducing individuals, the trait variance of clonally reproducing organisms (such as viruses) can be partitioned into variance components. In viral infections, only the virus genotype is passed on from one infection to the next, and thus the heritability of a viral trait as defined by h^2 can be used to quantify the relative influence of the variation in transmissible viral factors to the trait of interest in a population [8,20]. While the general framework of partitioning trait variance is the same in sexual and clonal populations, the underlying processes that generated the observed variance are different. Thus, although the methods for estimating heritability in sexual populations can generally be directly applied to viral heritability, care must be taken when evaluating the applicability of the methods.

Furthermore, whenever heritability is estimated, it is also important to remember that even though h^2 can be used to characterize the propensity of virus genetics to influence a trait, the actual measured heritability depends on both the expressed variation of this trait in a given population, as well as the variation in other factors that influence the trait. Therefore, h^2 strongly depends on the population it is measured in, and can consequently vary between study populations. This does not invalidate the usefulness of measuring h^2, but rather means that predictions on the basis of h^2 are primarily valid in the study population at hand, and care must be taken when using h^2 for predictions in other populations.

Parent–Offspring Regression

Parent–offspring regression is tightly linked to the original definition of heritability and is thus the most straightforward estimation method [17,18]. The idea is to compare trait values in parents to trait values in their respective offspring, or in the case of viral infections, compare traits from

Glossary

Asymptomatic phase: HIV infections are generally split into three characteristic stages: (i) primary infection/acute phase; (ii) chronic asymptomatic phase; (iii) AIDS phase.
Donor–recipient regression (DR): a method for estimating heritability by regressing the trait values in the recipients on the traits values of the donors.
Environmental variance: the amount of variance in the trait (e.g., SPVL) that is due to anything other then viral genetics.
Genetic bottleneck: a sudden decrease in population size where only a few genetic variants are selected. This leads to a drastic decrease in genetic variation in the population.
Genetic drift: change in the genotype distribution over time due to the finite size of a population.
Genetic variance: the amount of variance in the trait (e.g., SPVL) that is due to differences in transmissible viral genetics.
Pedigree: ancestral tree linking parents to their offspring.
Serodiscordant couples: sexual partnerships where only a single individual is infected.
Seronegative: negative for HIV infection.
Set-point viral load (SPVL): the viral load during the asymptomatic phase fluctuates around a remarkably stable level, the set-point viral load.
Viral heritability: the fraction of phenotypic variance that is explained by transmissible genetic factors.
Viral load: the density of virus in the blood of a patient. It is a proxy for the amount of virus in the rest of the body.