Introduction

The severity of HIV-1 infection is thought to result from an interplay of factors in the host, the virus, and the environment (for instance the presence of co-infections). Much work has focused on resolving host genetic factors which contribute to virulence [1], while the possible role of viral genetic factors inherited with the virus and transmitted with infection remains largely unresolved. Differences between viral subtypes are uncontroversial: for example subtype D appears to be associated with faster rates of disease progression [2–8]. But whether or not differences exist between more closely related strains, within subtypes, has not been established.

The existence of heritable viral factors influencing disease progression, and their contribution relative to other factors, is of interest for at least two reasons. Firstly, such factors, if they exist, have implications for how the virus influences the course of infection within an infected person. Secondly, if viral factors exist which affect virulence and can be preserved from one infection to the next, then these factors will be subject to natural selection at the population level [9]. In this study we test for the existence of such factors by examining viral load set-points among transmitting couples. HIV-1 viral load set-point is a quantitative measure of viral RNA copies in peripheral blood during asymptomatic infection. Viral load set-point is commonly used as a surrogate measure of the virulence of an infection since it is negatively associated with the time to AIDS and death [10].

At present, limited evidence suggests that viral load set-point is regulated by viral factors, although to an extent this reflects a paucity of research on the topic. The existence of a specific recombinant form associated with high viral loads strongly
The importance of unravelling the role of host and viral factors is donor index cases and the recipients of their blood products [15]. In cases of natural infection, an early study demonstrated macaques [12]. Other experimental evidence includes differences between SIV strains in the outcome of experimental infection of circumstances [11], as does the demonstration of stable differences suggests that viral factors can play a role in at least some transmission.

It has been hypothesised that the observed distribution of viral load set-point within transmitting couples, and that population-level selection could act upon these viral factors.

suggestions that viral factors can play a role in at least some circumstances [11], as does the demonstration of stable differences between SIV strains in the outcome of experimental infection of macaques [12]. Other experimental evidence includes differences between closely related HIV-1 strains in competition experiments [13,14]. In cases of natural infection, an early study demonstrated a correlation between the time to AIDS among infected blood donor index cases and the recipients of their blood products [15].

The importance of unravelling the role of host and viral factors is illustrated by the strong correlation in viral loads which has been observed in mother-to-child infections [16], which could be attributed to a combination of host and viral factors.

Finally, and most convincingly, a study of 115 HIV transmitting heterosexual couples in Zambia showed that 19% of variance in viral loads could be explained by shared homologous virus between couples \( p = 0.03 \) [17]. This study suggested a role for viral factors in determining viral load, but has not been repeated.

It has been hypothesised that the observed distribution of viral load set-points could be the result of natural selection acting on viral factors in order to maximise opportunities for transmission [9]. This hypothesis arose from an epidemiological analysis of the quantitative dependence between viral load, infectiousness and the duration of asymptomatic infection. This study demonstrated that people with the most common viral load set-points are predicted to be the most productive in terms of onward transmission over the course of infection; lower viral loads are associated with a longer life expectancy and thus more opportunities for transmission, but this is offset by reduced infectiousness. Conversely, those with higher viral load set-points are more infectious, but progress to AIDS too quickly to produce as many onward infections over the whole course of their asymptomatic period. In other terms, the observed distribution of viral load set-point is consistent with an evolutionary life-history trade-off for the virus [9]. If this interpretation is correct, then the observed distribution of viral loads set-points and, by extension, virulence, could be the product of viral adaptation acting to maximise opportunities for transmission.

For this hypothesis to be correct, viral load set-point must be a heritable property, partly determined by the virus and preserved from one infection to the next. If it is not heritable, there is no way natural selection can act upon it. In this study we estimate heritability in viral load set-point within transmitting couples, and account for a number of important confounding factors. We estimate heritability as the proportion of variance in viral load set-point which is determined by infection with genetically similar virus for HIV-1-infected heterosexual couples identified in the Rakai District of south-western Uganda.

Methods

Study population

The study population was enrolled in the Rakai Community Cohort Study in the rural Rakai District of south-western Uganda. Study methods have been described in detail elsewhere [18,19], but are briefly outlined here.

More than 12,000 consenting subjects aged 15–49 were interviewed in surveys conducted at 10–12 month intervals from 1994–2003. Participants provided written, informed consent; and were provided with condoms and voluntary HIV counselling and testing free of charge. Participants agreed to provide identifying information for their married or consensual partners which allowed retrospective linkage of couples. The study was approved by review boards at the Uganda Virus Research Institute, the AIDS Research Subcommittee of the Ugandan National Council for Science and Technology, Columbia University, and Johns Hopkins University.

HIV prevalence in the cohort was 16.5%, and average annual HIV incidence was 1.5 cases/100 person-years [19].

Retrospective analyses identified 200 self-reporting sexual partners for whom there was evidence of seroconversion for one or both partners during the course of the study. The partner who was seropositive first was identified as the index case, and the other partner as the secondary transmission case. For some couples the ordering of events could not be identified because they both seroconverted within the same round of the study. For some of these concurrently infected couples, the partner reporting an external sexual relationship could be inferred to be the index individual.

Serum samples from venous blood provided at survey visits were tested for HIV-1 RNA levels quantified by a reverse-transcriptase polymerase chain reaction (RT-PCR) assay (Amplicor HIV-1 Monitor 1.5 assay, Roche Molecular Systems) with a lower detection limit of 400 copies/mL \((2.6 \log_{10} \text{copies/mL})\). Antiretroviral therapy (ART) was not available in Rakai at the time of the study, but participants were offered free general health care and treatment for opportunistic infections.

During early infection and prior to AIDS and death, viral loads are elevated above the set-point. To exclude data from early infection, viral loads measured at the first visit with a positive serology following a previous visit with negative serology were excluded. To ensure measurements made during late infection were also excluded, viral loads from the last observation prior to death \((\text{up to a maximum of 12 months prior to death})\) were also excluded. Following these exclusions 112 couples were identified for whom suitable viral load measurements were available. For those individuals with more than one viral load measurement, the set-point was defined as the mean \(\log_{10}\) viral load over eligible visits. Age at the time of measurement was also averaged.

HIV-1 subtype was determined for 171 individuals using a Multi-region Hybridization Assay (MHAacd) on serum samples [20] as previously described for this cohort [6,21]. Samples were classified as subtype A, D, C and A/D recombinants. 8 couples
whose subtypes were discordant (where subtypes were available for both individuals) were excluded from the statistical analysis.

For a subset of the couples viral sequence data from the gag (p24) and gp41 regions were available for comparison in both partners to help identify transmitting couples. To decrease the risk of spurious linkages between sequences, p24 and gp41 sequences for the couples were analysed together with sequences from 511 other infected individuals in the cohort [22]; in total 620 p24 and 614 gp41 sequences were analysed (See Text S1 for Genbank accession numbers). For 603 of these, sequences at both loci were available for a particular individual. In these cases a phylogenetic analysis was conducted on the concatenate of the two sequences. A European subtype B virus, accession number EU786678.1, was used as the outgroup for all loci. The sequences are approximately 400 base pairs long, which is sufficient to cluster sequences for our purposes. Phylogenies were derived by maximum likelihood methods using a genetic substitution model chosen among many to best represent the data. The most appropriate substitution model was selected by comparing the rapid maximum likelihood fits in jModelTest v.0.1.1 [23,24] by Akaike Information Criteria (AIC). The model selected was a general time reversible (GTR) nucleotide substitution model with a gamma distribution of rates (+G) and a proportion of invariant sites (+I) was used. The GTR+G+I model was the most suitable model among 85 candidate models for the concatenated sequences and gp41 sequences by the Akaike Information Criterion (AIC). It was the third most appropriate model for p24 (with AIC = 12.2) but was also used for this locus for comparability between loci.

The phylogenetic analysis was conducted using RAxML 7.0.3 [25] which produced a maximum likelihood tree using a rapid bootstrap algorithm (100 replicates) [26]. The bootstrap values written on to this tree were determined by a further 1000 bootstrap replicates produced by a rapid hill-climbing algorithm [27].

Concatenates of both loci were used to assess phylogenetic support for epidemiological linkage where they were available for both individuals (36 couples). For 14 couples, sequence data were only available for both individuals at p24, and for 8 couples data were only available for both individuals from gp41. In these cases it was only possible to perform the analysis based on single loci.

Couples were considered to be strongly linked if their sequences were monophyletic and the clade had bootstrap support of greater than or equal to 80%. This condition was imposed on both single locus and concatenated sequences to account for the possible effects of recombination in distorting the phylogenetic signal. Our approach to determining linkage within couples is thus conservative. Couples were considered to have no support for linkage if their sequences were polyphyletic.

To analyze the data on viral load set-points within transmitting couples, we performed our analysis on two groups. The first group included all couples, with and without genetic data, but excluding both those with sequence data who had no genetic support for linkage and those with discordant subtypes determined by MHAxoo. The second subgroup included couples with strong genetic support for linkage (monophyletic with greater than 80% bootstrap support).

Symptoms of genital ulcer disease (GUD) over the interval prior to sample collection were ascertained via interview, and by physical examination for ulcers reported to be present at the time of a study visit. GUD has previously been found to be a significant predictor of viral load in this cohort [29]. If either or both partners had GUD which raised their viral loads during the study period this might confound the correlation of viral loads between individuals within couples. A report of any GUD in the six months prior to or at the time of viral load measurements was considered to be presence of GUD.

Since the data used in this study were not collected with the analysis presented here in mind, there are incomplete data on sequencing and epidemiological data. This may lead to unidentified biases in the data.

Statistical analysis

We used analysis of variance (ANOVA) to test whether there was greater similarity in viral load set-points of individuals within transmitting couples than between all individuals. In other words, we decomposed the variance in viral load set-point into the sum of within-couple variance and between-couple variance. To perform ANOVA, a general linear model was formulated with a regression coefficient for each couple (see Text S2). In a first unadjusted analysis, the best estimate of these regression coefficients is the mean of the viral load set-points of two individuals in a couple. The significance test is then a comparison of this model (where each viral load is predicted by the coefficient for the couple) versus the null model (where there is only one coefficient, the overall mean for all individuals).

The strength of the effect is measured by $R^2$, the proportion of variance explained by the model. Since one parameter is introduced for each couple, a proportion of variance is explained spuriously due to decreased residual degrees of freedom. The adjusted $R^2$, denoted $R^2_a$, is defined as the proportion of the remaining variance explained and accounts for this spurious effect.

To adjust for possible confounders, the general linear model was extended to include the effects of gender, age and GUD status, which have all been previously shown to affect viral load set-point within this study population [28], and role in transmission (index or secondary case).

Since the biological origin of any similarity in viral loads is hypothesized to be due to similarity in viral genotypes between transmitting individuals, the measured association may be interpreted as an estimate of the effect of viral genotype on viral load set-point. In this context, the study design is analogous to pedigree studies in classical genetics which are used to study the association between genotype and phenotype [29]. Broad-sense heritability is defined as the ratio of genotypic variance to phenotypic variance. In our study, heritability is estimated by the ratio of variance in viral load set-points within transmitting couples, to variance in viral load set-points in the population as a whole [29]. In other words, heritability and $R^2_a$ are equivalent concepts.

We thus estimate heritability, as $R^2_a$ for the single factor model. The estimate can further be adjusted for confounders, denoted here by $R^2_a$ [see Text S2 and [30]].

The validity of our statistical approach is supported by the observation that the $p$-values obtained from the unadjusted analysis were equal to the proportion of permutation tests (repeatedly sampling and re-linking individuals into random pseudo-couples) which gave the same or larger $R^2_a$, and also to $p$-values obtained by comparing the distribution of differences in viral loads within and between couples, thus confirming the validity of ANOVA to analyse these data (analysis not shown).

A related question of interest is the extent to which the viral load set-point of one individual can be used to predict the set-point of the person they infect. The strength of association in a unidirectional analysis (the correlation coefficient, $\rho$) is equal to heritability, a relation which can be shown to hold exactly for viral loads distributed according to a bivariate Normal distribution, and also holds for the data analysed here (not shown) [29].
Figure 1. Clustering of sequences from the couples for whom sequence data was available. Sequences from the couples were analysed together with sequences from 511 other infected individuals in the cohort ('filler sequences') to prevent spurious linkage due to independent infections with similar circulating virus. For the sake of clarity, filler sequences are not shown in this figure (the full trees are shown in Figure S1, S2 and S3). Sequences from couples are categorised as polyphyletic (red), monophyletic with bootstrap >80% (blue) or monophyletic with bootstrap <80% (red).
As outlined above, we performed this analysis on the large group of couples with moderate support for transmission and a subset with strong genetic support for transmission. The first group included all couples for whom there was epidemiological linkage and, where data was available, at least moderate genetic support for transmission. The second, more conservative, subgroup included only those with strong genetic support for transmission. We were thus able to investigate whether the signal became stronger when stricter inclusion criteria were imposed.

Results

Phylogenetic analysis

The phylogenetic trees used to identify the level of linkage between couples are shown in Figure 1. The additional sequences included to prevent spurious linkage have been excluded from the figure for clarity (full trees are shown in Figure S1, S2 and S3). The outcome of the phylogenetic clustering analysis and resulting inclusion criteria for the ANOVA are summarised in a flow chart (Figure S4).

Of the 35 couples with data available for both individuals at both p24 and gp41, 31 were monophyletic, with 29 showing greater than 80% bootstrap support. The remaining 4 couples were polyphyletic.

14 couples had sequence data available for both partners at p24 alone. Of these, 12 were monophyletic, 7 of which had greater than 80% bootstrap support, and 2 were polyphyletic (Figure 1B). Of the 8 couples with data available for both partners at gp41 alone, 7 were monophyletic, with 6 showing greater than 80% bootstrap support (Figure 1C).

Overall, 29 of the 57 couples with viral sequence data showed strong support for intra-couple transmission based on genetic linkage (monophyletic with >80% bootstrap support on a single locus or multiple loci where available). There were indications that 8 couples (14%) did not transmit to each other and the remaining 21 couples were indeterminate (37%). The couples with strong support for transmission have distinctly closer tree distances than the rest of the sample (Figure S5).

Table 1. Mean log_{10} HIV load set-point.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Couples with moderate support for transmission (97 couples)</th>
<th>Subgroup of couples with strong support for transmission (29 couples)</th>
</tr>
</thead>
<tbody>
<tr>
<td>HIV load, mean log_{10} copies/mL (SD)</td>
<td>HIV load, mean log_{10} copies/mL (SD)</td>
<td></td>
</tr>
<tr>
<td>All</td>
<td>4.39 (0.84)</td>
<td>58 (100%)</td>
</tr>
<tr>
<td>Gender</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>4.42 (0.85)</td>
<td>29 (50%)</td>
</tr>
<tr>
<td>Female</td>
<td>4.36 (0.84)</td>
<td>29 (50%)</td>
</tr>
<tr>
<td>Age, years</td>
<td></td>
<td></td>
</tr>
<tr>
<td>15–24</td>
<td>4.45 (0.74)</td>
<td>14 (24%)</td>
</tr>
<tr>
<td>25–29</td>
<td>4.40 (0.77)</td>
<td>20 (34%)</td>
</tr>
<tr>
<td>30–39</td>
<td>4.59 (0.70)</td>
<td>18 (31%)</td>
</tr>
<tr>
<td>40–64</td>
<td>4.63 (0.80)</td>
<td>4 (7%)</td>
</tr>
<tr>
<td>Missing</td>
<td>3.52 (1.02)</td>
<td>2 (3%)</td>
</tr>
<tr>
<td>GUD</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Present</td>
<td>4.52 (0.73)</td>
<td>14 (24%)</td>
</tr>
<tr>
<td>Absent</td>
<td>4.57 (0.78)</td>
<td>39 (67%)</td>
</tr>
<tr>
<td>Missing</td>
<td>3.87 (0.87)</td>
<td>5 (9%)</td>
</tr>
<tr>
<td>Subtype</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>4.52 (0.67)</td>
<td>10 (17%)</td>
</tr>
<tr>
<td>C</td>
<td>1 (1%)</td>
<td></td>
</tr>
<tr>
<td>D</td>
<td>4.53 (0.76)</td>
<td>36 (62%)</td>
</tr>
<tr>
<td>Recombinant</td>
<td>4.37 (0.81)</td>
<td>12 (21%)</td>
</tr>
<tr>
<td>Missing</td>
<td>3.36 (0.95)</td>
<td></td>
</tr>
<tr>
<td>Transmission</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Index partner</td>
<td>4.52 (0.72)</td>
<td>23 (40%)</td>
</tr>
<tr>
<td>Secondary case</td>
<td>4.35 (0.87)</td>
<td>23 (40%)</td>
</tr>
<tr>
<td>Missing</td>
<td>4.24 (0.97)</td>
<td>12 (21%)</td>
</tr>
</tbody>
</table>

Viral load set-point for individuals in the 97 seroconverting couples with moderate support for transmission and in the subgroup of 29 couples with strong genetic support for transmission by sex, age, genital ulcer disease (GUD) status, subtype and role in transmission.

doi:10.1371/journal.ppat.1000876.t001
Following the phylogenetic and subtype analysis, the statistical analyses were performed on 97 couples with moderate support for transmission (Figure S4) and a subgroup of 29 couples who had strong support for transmission.

**Viral loads**

The average log_{10} RNA viral load set-point was 4.39 log_{10} cps/mL, with values in the range 2.60 log_{10} cps/mL (the limit of detection) to 7.14 log_{10} cps/mL (Table 1). The average duration of follow up was just under a year (352 days) and 3 viral load datapoints from which to calculate set-point. Nearly half of individuals had only one valid viral load datapoint (86 of 194, 44%). The majority of couples were infected with subtype D viruses (63% of all individuals), with subtype A the second most common subtype (12%) (Table 1). Amongst individuals for whom GUD status was known, the majority were GUD negative, across all groups, and no significant association was found between GUD within couples ($p = 0.14$).

Viral load set-points of transmitting couples are presented in Figure 2. This data is also presented as the distribution of differences in viral load set-points, Figure S6.

When analysing the 97 couples with moderate support for transmission, the couple effect (which tests whether viral load set-

![Figure 2](#)

**Figure 2. Viral load set-point of index partner versus that of the secondary case in transmitting couples.** Couples are stratified by male to female transmission (green triangles), male to female transmission (blue circles) and unknown direction of transmission (red diamonds, plotted as female against male viral load, since the index partner could not be identified). **A** Couples with moderate support for transmission ($n = 97$). **B** Subgroup of couples with strong genetic support for transmission ($n = 29$, monophyletic and bootstrap ≥80%). Simple linear regression lines are not shown since this was not the analysis performed.

doi:10.1371/journal.ppat.1000876.g002
points are similar within couples, see Methods) was found to be borderline significant ($p = 0.054$) by single factor ANOVA. The estimated size of this couple effect was 16%.

In addition, age, GUD and subtype were found to be highly significant (Table 2) in accordance with earlier studies of this cohort [28]. An unexpected finding is that individuals with missing data generally had lower viral loads than other individuals, which suggests there may be some selection bias not captured in this study (Table 1). For this reason, we treated ‘missing data’ as a separate categorical state for the corresponding variables. When the single factor models were instead fitted excluding the missing data, age, GUD and subtype were not found to be significant predictors of viral load.

When adjusting for all possible confounders (in a multivariate ANOVA), the couple effect was a significant predictor of viral load set-point (adjusted $p = 0.0059$ for full model). The adjusted estimate for the couple effect was 23% (Table 2). To test the robustness of our conclusions to inclusion of different confounders, we explored all possible combinations of factors (Table S1).

When looking at the subgroup of couples showing strong support for transmission (29 couples), the couple effect was of borderline significance in the unadjusted analysis ($p = 0.067$), but significant when adjusting for confounders ($p = 0.036$). The size of the couple effect was 27%. The couple effect was a key determinant of viral load for most of the multivariate analyses which were performed (Table S2). When adjusting for confounders the estimate of the couple effect increased to 37% (Table 2).

The set of 15 couples for whom there was no support for transmission (i.e. with different serotypes or polyphyletic viral genotypes) might be considered as a small control group for our study. Unfortunately, this group is too small to form a definitive

Table 2. Factors which influence viral load set-point.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient (SE)</th>
<th>$p$</th>
<th>$R^2$</th>
<th>Coefficient (SE)</th>
<th>$p$</th>
<th>$R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Single factor models</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Couple</td>
<td>0.054</td>
<td>0.16%</td>
<td></td>
<td>0.067</td>
<td>0.27%</td>
<td></td>
</tr>
<tr>
<td>Sex</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>0.068 (0.12)</td>
<td>0.58</td>
<td>-0.4%</td>
<td>-0.112 (0.21)</td>
<td>0.92</td>
<td>-1.3%</td>
</tr>
<tr>
<td>Female</td>
<td>-</td>
<td>-</td>
<td></td>
<td></td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Age, years</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>40–64</td>
<td>1.11 (0.22)</td>
<td>&lt;0.0001</td>
<td>14%</td>
<td>0.80 (0.67)</td>
<td>0.91</td>
<td>4.7%</td>
</tr>
<tr>
<td>30–39</td>
<td>1.08 (0.20)</td>
<td>&lt;0.0001</td>
<td>14%</td>
<td>1.36 (0.57)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>25–29</td>
<td>0.88 (0.20)</td>
<td></td>
<td></td>
<td>1.07 (0.57)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>15–24</td>
<td>0.94 (0.21)</td>
<td></td>
<td></td>
<td>1.04 (0.58)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GUD</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Present</td>
<td>0.65 (0.19)</td>
<td>&lt;0.0001</td>
<td>11%</td>
<td>0.95 (0.40)</td>
<td>0.39</td>
<td>6.58%</td>
</tr>
<tr>
<td>Absent</td>
<td>0.70 (0.14)</td>
<td></td>
<td></td>
<td>0.81 (0.36)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Missing</td>
<td>-</td>
<td>-</td>
<td></td>
<td></td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Subtype</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>1.15 (0.24)</td>
<td>&lt;0.0001</td>
<td>15%</td>
<td>0.37 (0.33)</td>
<td>0.19</td>
<td>3.2%</td>
</tr>
<tr>
<td>C</td>
<td>1.33 (0.80)</td>
<td></td>
<td></td>
<td>None</td>
<td></td>
<td></td>
</tr>
<tr>
<td>D</td>
<td>1.16 (0.19)</td>
<td></td>
<td></td>
<td>0.51 (0.26)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Recombinant</td>
<td>1.01 (0.23)</td>
<td></td>
<td></td>
<td>1.01 (0.23)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Missing</td>
<td>-</td>
<td>-</td>
<td></td>
<td></td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Role in transmission for transmitting pairs</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Index case</td>
<td>0.29 (0.16)</td>
<td>0.17</td>
<td>0.8%</td>
<td>0.46 (0.28)</td>
<td>0.82</td>
<td>2.2%</td>
</tr>
<tr>
<td>Secondary case</td>
<td>0.12 (0.16)</td>
<td>0.45 (0.28)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unknown</td>
<td>-</td>
<td>-</td>
<td></td>
<td></td>
<td>-</td>
<td></td>
</tr>
<tr>
<td><strong>Multiple factor models</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>All factors</td>
<td>0.0057</td>
<td>0.40%</td>
<td></td>
<td>0.036</td>
<td>0.40%</td>
<td></td>
</tr>
<tr>
<td>All factors except couple</td>
<td>0.17%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$R^2$</td>
<td>23%</td>
<td></td>
<td></td>
<td>37%</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Individuals were stratified by sex, age, genital ulcer disease (GUD), viral subtype and paired within couples. Regression coefficients, $p$-values and $R^2$ adjusted for degrees of freedom, $R^2_{adj}$ were reported for single factor models. For the model including all factors the $p$-value of the couple effect (see methods) and $R^2$ for the couple effect adjusted for confounders, $R^2_{adj}$, are presented (see Text S2). The analysis was performed for all 97 couples with moderate support for transmission and for a subset of 29 couples with strong support for genetic linkage.

doi:10.1371/journal.ppat.1000876.t002
control group. Nonetheless, for completeness, we estimated the couple effect. It was not found to be significant ($p = 0.32$) and the effect size was smaller (12%).

**Discussion**

Our analysis showed that, in this study population, individuals within transmitting couples had similar viral load set-points ($p = 0.054$ in single factor model, $p = 0.0057$ adjusting for confounders) and that this effect explained 16% (23% adjusting for confounders) of the variability in viral load set-points. When the analysis was repeated for the subgroup of couples for whom there was strong genetic support for viral linkage, couples infected with similar viruses also had similar viral load set-points ($p = 0.067$, adjusted $p = 0.036$). The size of the couple effect was estimated to be larger, 27% in single factor model (37% adjusting for confounders), suggesting that the transmitted virus plays a role in determining viral load set-point.

We were unable to assess and account for all possible contributing factors to the correlation of viral loads within couples. Potential confounders include environmental or host factors which could cause couples to have similar viral load set-points. For example, couples may have similar exposure to coinfections or access to health care which might affect viral load set-point.

Besides environmental factors, the viral load set-points of secondary cases could depend on the ‘dose’ of transmitted virus received from the index case. If the dose were to depend on the set-point of the index case, this would lead to correlated viral load set-points. Phylogenetic analysis of 102 early infection isolates indicated that 78 of these infections were established by a single virus, and that the remaining 24 were established by two to five viruses [31]. The viral load set-point of the index partner was not known for that study and therefore the relationship between dose and number of viruses establishing infection is not known. In addition, the relationship between the number of establishing virions and the viral load set-point of the recipient partner is not known. However, since most infections were established by only a few virions, or resulted from the rapid outgrowth of the population descended from these virions, it is likely that the number of infecting virions is similar for a large range of viral load set-points of the infecting partner. Given all these unknown relationships, the hypothesis that a dose effect is driving the observations presented here cannot be discounted, and may be further elucidated by ongoing study in humans and experimental infections of animals [32–34].

The most parsimonious explanation for our observation is the existence of viral virulence factors that influence viral load set-point and are partly preserved from one infection to the next. The existence or identity of these viral factors is not well established. Candidate virulence factors include the accumulation of CTL escape mutations at a population level [35], traits determined by viruses preserved on mucosal surfaces by balancing selection [36], and other virulence factors acting by presently unknown mechanisms.

This retrospective study of heterosexual couples in a rural African population suggests that the transmitted virus plays an important role in determining viral load set-point, supporting previous observations [17]. Our study is likely to give an underestimate of the role of viral factors in determining viral load for three main reasons. The infecting viruses in almost all these couples were not identical, only similar, there were only a few viral load measurements per individual and so variability within patients could not be accounted for and we had no information on the host genetics of the infected individuals. Remaining variability in viral load set-point could be due to various host immune factors, coinfections and other environmental factors. The suggestion that the virus plays a role in determining viral load set-point should not negate the importance of host factors [1,37], but rather implies a complex interaction between host and virus.

The similarity of viral load set-points between transmitting couples, as demonstrated in our analysis, have direct implications for potential of HIV-1 virulence to evolve both in untreated infection and in response to public health measures [9]. More extensive studies with greater numbers of couples, more detailed virus and host genetic data and different routes of transmission are required to further test our observation.

**Supporting Information**

**Text S1** Accession numbers
Found at: doi:10.1371/journal.ppat.1000876.s001 (0.29 MB PDF)

**Text S2** Statistical model and partitioning variance
Found at: doi:10.1371/journal.ppat.1000876.s002 (0.39 MB PDF)

**Figure S1** Clustering of sequences from the couples for whom sequence data was available at both loci, based on concatenated sequences. As Figure 1A, but with all sequences shown.
Found at: doi:10.1371/journal.ppat.1000876.s003 (0.75 MB PDF)

**Figure S2** Clustering of sequences from the couples for whom sequence data was available at gp120. As Figure 1B, but with all sequences shown.
Found at: doi:10.1371/journal.ppat.1000876.s004 (0.76 MB PDF)

**Figure S3** Clustering of sequences from the couples for whom sequence data was available at p24. As Figure 1C, but with all sequences shown.
Found at: doi:10.1371/journal.ppat.1000876.s005 (0.76 MB PDF)

**Figure S4** Flow diagram for inclusion in the study groups of 97 couples with moderate support for transmission, which includes those with epidemiological linkage together and where available, weak to strong support for transmission (blue); the sub-group of 29 couples with strong support for transmission (green); the 15 couples which genetic evidence suggested did not transmit to each other (red).
Found at: doi:10.1371/journal.ppat.1000876.s006 (0.58 MB PDF)

**Figure S5** Distribution of tree distances. The distribution of tree distances between couples is given for couples for whom there was strong support for transmission (green) and weak genetic support for transmission (blue). In addition, the distribution of tree distances for all other pairwise comparisons between individuals in the trees (Figure 1) is included for comparison (black). For couples with sequences at both loci available the distance shown is that on the concatenated tree. For couples for whom sequence data was only available at one locus, the distance on that single locus tree is used.
Found at: doi:10.1371/journal.ppat.1000876.s007 (0.01 MB PDF)

**Figure S6** Distribution of differences in viral load set-points. The distribution of absolute differences in viral load setpoints for 29 couples with strong support for transmission (green), the remaining 68 couples from the 97 with moderate support for transmission (blue) and all other male to female pairwise comparisons (black).
Found at: doi:10.1371/journal.ppat.1000876.s008 (0.08 MB PDF)

**Table S1** Size of couple effect for different model structures for 97 couples in main analysis. Black circles indicate factors included.
in the model. The Type III p-value for the couple effect and the adjusted R-squared for the model are given.

Table S2  Size of couple effect for different model structures for subgroup of 29 couples. Black circles indicate factors included in the model. The Type III p-value for the couple effect and the adjusted R-squared for the model are given.

References


Acknowledgments

We thank Nicholas Grassey, Andrew Leigh Brown, Benjamin Bolker, Alexei Drummond and Andrew Rambaut for useful suggestions. We are also grateful to anonymous reviewers for helpful input.

Author Contributions

Conceived and designed the experiments: TDH CF. Performed the experiments: OL DS MJW NK FN ACS VS TCG RHG. Analyzed the data: TDH OL GS CAD ACS WPH CF. Wrote the paper: TDH OL WPH TCG RHG CF.

Similar Viral Loads in HIV-1 Transmitting Couples