OPEN ACCESS Freely available online

PLOS | PATHOGENS

The Contribution of Viral Genotype to Plasma Viral Set-Point in HIV Infection

Emma Hodcroft[1], Jarrod D. Hadfield[1], Esther Fearnhill[2], Andrew Phillips[3], David Dunn[2], Siobhan O'Shea[4], Deenan Pillay[5], Andrew J. Leigh Brown[1]*, on behalf of the UK HIV Drug Resistance Database and the UK CHIC Study[¶]

1 Institute of Evolutionary Biology, University of Edinburgh, Ashworth Laboratories, Edinburgh, United Kingdom, **2** MRC Clinical Trials Unit Aviation House, London, United Kingdom, **3** Infection and Population Health, University College London, Royal Free Hospital, London, United Kingdom, **4** Department of Infectious Diseases, King's College London, London, United Kingdom, **5** Research Department of Infection, University College London, London, United Kingdom

Abstract

Disease progression in HIV-infected individuals varies greatly, and while the environmental and host factors influencing this variation have been widely investigated, the viral contribution to variation in set-point viral load, a predictor of disease progression, is less clear. Previous studies, using transmission-pairs and analysis of phylogenetic signal in small numbers of individuals, have produced a wide range of viral genetic effect estimates. Here we present a novel application of a population-scale method based in quantitative genetics to estimate the viral genetic effect on set-point viral load in the UK subtype B HIV-1 epidemic, based on a very large data set. Analyzing the initial viral load and associated *pol* sequence, both taken before anti-retroviral therapy, of 8,483 patients, we estimate the proportion of variance in viral load explained by viral genetic effects to be 5.7% (CI 2.8–8.6%). We also estimated the change in viral load over time due to selection on the virus and environmental effects to be a decline of 0.05 $\log_{10}$ copies/mL/year, in contrast to recent studies which suggested a reported small increase in viral load over the last 20 years might be due to evolutionary changes in the virus. Our results suggest that in the UK epidemic, subtype B has a small but significant viral genetic effect on viral load. By allowing the analysis of large sample sizes, we expect our approach to be applicable to the estimation of the genetic contribution to traits in many organisms.

Citation: Hodcroft E, Hadfield JD, Fearnhill E, Phillips A, Dunn D, et al. (2014) The Contribution of Viral Genotype to Plasma Viral Set-Point in HIV Infection. PLoS Pathog 10(5): e1004112. doi:10.1371/journal.ppat.1004112

Editor: Michael Worobey, University of Arizona, United States of America

Received September 24, 2013; **Accepted** March 22, 2014; **Published** May 1, 2014

Funding: This work was supported by the Biological Sciences Research Council studentship (www.bbsrc.ac.uk) and funds from the Royal Society (royalsociety.org). The UK HIV Drug Resistance Database is supported by the Medical Research Council (grant number G0900274) (www.mrc.ac.uk) and is partly funded by the Department of Health (www.gov.uk/government/organisations/department-of-health). Additional support for the UK HIV RDB is provided by Boehringer Ingelheim (www.boehringer-ingelheim.co.uk), Bristol-Myers Squibb (www.b-ms.co.uk), Gilead (www.gilead.com), Tibotec (a division of Janssen-Cilag) (www.janssentherapeutics.com) and Roche (www.roche.co.uk). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: Additional support for the UK HIV Resistance Database is provided by Boehringer Ingelheim, Bristol-Myers Squibb, Gilead, Tibotec (a division of Janssen-Cilag) and Roche. Although some authors have received funding from various commercial organizations for research, travel grants, speaking engagements or consultancy fees, neither those organizations nor the study funders had any role in study design, data collection and analysis, decision to publish, or preparation of the manuscript and this does not alter our adherence to all PLOS Pathogens policies on sharing data and materials.

* E-mail: A.Leigh-Brown@ed.ac.uk

¶ Membership of the UK HIV Drug Resistance Database and the UK CHIC Study is provided at the end of this paper.

Introduction

Plasma viral load has long been considered one of the most important clinical measures in HIV-positive patients. The progression time from infection to AIDS or death varies enormously from a few years to decades, and 'set-point' viral load, taken early in the asymptomatic phase of the disease, is the best known early predictor of the long-term rate of disease progression [1–3] and is also strongly associated with transmission risk [4–6]. Variation in host genes, particularly HLA [7–12] but also the CCR5 co-receptor and its ligands, and even the gene APOBEC3G [13–15], has been identified as influencing progression rate, but the contribution of the viral genome is still much less clear.

Nevertheless, the hypothesis that HIV could be evolving to become more virulent has been a driver for decades of HIV research. In the mid-1980's, it became clear that some HIV isolates, deemed 'high/fast' lines, had a much higher replicative capacity in cell lines than others [16–18]. When a drop in CD4+ cell count at diagnosis was reported a few years later [19,20], speculation began as to whether the spread of these 'high/fast' lines could be responsible [20–22]. A number of studies looking at long-term trends in HIV virulence were published, drawing mixed conclusions on whether there was evidence of HIV becoming more virulent [23–35]. However, a lack of standardization of when measurements were taken, what measures were used, and whether patients were on anti-retroviral therapy (ART), as well as differences in the subtypes, risk groups, and demographics of the patients involved mean that these studies are difficult to compare directly. Despite this, two meta-analyses have been performed, both concluding that a decrease in CD4+ count and an increase in viral load can be observed, implying an increase in HIV virulence

Author Summary

HIV viral load, the amount of virus in the blood, is an important predictor of rate of CD4+ cell decline, time to AIDS and onwards transmission. Plasma viral load is influenced by many environmental and host factors, but the contribution of the viral genome is not yet clear. We have adapted a method from quantitative genetics which considers the viral phylogeny as a pedigree, permitting analysis of large cohort-derived datasets for the first time. We found the viral genome contributes significantly to the level of the set point viral load, but only determines about 6% of the variation in this property in this population. Our study also suggests that the change over time in mean plasma viral load described in some recent studies has not been due to a change in the component of viral load that is contributed by viral genotype.

that both papers suggest could be caused by viral factors [36,37]. This would require that the viral genome exerted some influence over the set-point viral load. In the context of drug resistance it is well known that viral variation affects the replication capacity of HIV (reviewed in [38]), suggesting that such a viral genetic influence could indeed be possible.

Evolutionary theory predicts that pathogens evolve to modulate their density within hosts in order to maximize transmission rate. In the classic studies of myxomatosis [39], viral genotypes with reduced replication rates that permitted longer host survival were selected for when host density, and thus transmission probability, declined as the epidemic progressed. This, along with classic studies on the link between transmission and virulence [40,41], raises the possibility that in the 100 years HIV is known to have infected humans [42,43], it might have adapted to different levels of transmission probability associated with different infected populations [2,6]. Studies of disease progression and viral load have found evidence of differences between HIV-1 subtypes [44–46], suggesting that major viral genetic differences among immunodeficiency viruses influence virulence.

Three studies investigated the contribution of viral genotype to set-point in studies of HIV sero-discordant couples. In these studies of 115, 56, and 47 sequence verified transmission-pairs in Zambia, the Netherlands and the USA, correlation coefficients of 0.21, 0.25, and 0.55, respectively, were estimated between set-point viral load in the index and contact cases [10,47,48]. Another transmission-pair study on 28 couples from Uganda reported the coefficient of determination from ANOVA analysis as $R^2 = 27\%$, and $R^2 = 37\%$ after adjusting for confounding effects [49]. In a fourth study, based on the Swiss Cohort, Alizon et al. [50] adopted a phylogenetic approach, looking for a signal of inherited viral effect in men who have sex with men (MSM) infected with subtype B. Phylogenetic signal measures the amount that the connections in a phylogeny explain the similarity in trait values seen in different individuals [51,52]. Using their strictest definition of set-point viral load and consequently their smallest sample size (n = 134), Alizon et al. [50] obtained a statistically significant estimate that approximately half of the variation observed in viral load could be explained by viral genetic effects. However, the estimates obtained using a more liberal definition of set-point viral load in the MSM group (n = 404) were much lower, at around 11%, and in the largest datasets where all risk-groups were included the estimates were non-significant.

Given the small numbers in all of these studies, we sought an alternative approach which would allow the inclusion of the large numbers of individuals for which both plasma viral load and viral sequence data are now available.

In quantitative genetics the proportion of the total trait variation (V_P) caused by additive genetic factors (V_A) is described as its narrow-sense heritability (h^2). Numerous approaches have been proposed to estimate variance components and heritability from phylogenetic data, including restricted maximum-likelihood (REML) [53], maximum-likelihood (ML) [51], and generalized least squares [52]. REML methods have emerged as the preferred choice for variance component and heritability estimation due to their ability to give unbiased estimates [54–56]. In 1996 the program ASReml introduced an efficient implementation of REML-based variance estimation specifically designed for data from pedigreed individuals [56,57]. By measuring the relationships between individuals on the pedigree as the probability that their alleles are identical by descent, and linking this to the observed differences in trait measures, the amount of trait variation explained by the genetic relationships can be estimated. These identical by descent relationship measures are calculated from the pedigree to form a genetic relatedness matrix, usually referred to as $\mathbf{A}$ [58].

For a phylogeny, the phylogenetic covariance of two taxa is proportional to the total length from the taxa's most recent common ancestor (MRCA) to the root under a Brownian motion model of evolution [59,60], and the covariances between all taxa can be represented by the matrix $\mathbf{A}$. In order to calculate variance components the inverse of $\mathbf{A}$, $\mathbf{A}^{-1}$, is usually needed, but can be computationally resource intensive to calculate [56,58]. Henderson [58] showed that for pedigrees this problem can be made easier by including 'phantom parents' for all individuals with unknown parentage so that the population could be traced back to unrelated ancestors. Hadfield and Nakagawa [56] extended this technique to phylogenies by expanding $\mathbf{A}$ to include all the internal nodes in the tree, allowing the inverse matrix to be calculated by the method of Henderson [58] and to provide a structure to the model that can be exploited by generic sparse matrix algorithms [61]. (See [62] for an alternative algorithm.)

Here, we apply this approach by using ASReml to estimate the heritability of viral load in the UK subtype B HIV epidemic, analyzing set-point viral load in almost 8,500 individuals for whom matched HIV sequences and viral load data were available.

Results

The sequences used were made available by the UK HIV Drug Resistance Database (UK HIV RDB), which collects *pol* sequences from HIV-positive patients attending clinics across the UK before starting and during ART in order to detect resistance mutations. The UK HIV RDB was estimated to contain sequences for approximately two-thirds of the subtype B MSM patients who were treated for HIV in the UK in 2006 [63]. The first sequence available for each patient was analyzed. Fully anonymized clinical data corresponding to many of the sequences was made available by the UK Clinical HIV Cohort (UK CHIC) [64], with viral load before starting ART being available for 8,700 initial subtype B sequences, reflecting the most prevalent subtype epidemic in the UK. The data used were the most current available, with sequences and clinical data collected up to mid-2009.

After removing all cases where there was uncertainty over disease or treatment status or large sections of sequence were missing, 8,483 subtype B sequences and associated viral load measurements remained. The demographics of the dataset show that 73% (6,198 individuals) were white MSM, reflecting the historic preponderance of this subtype among MSM (Table 1). A phylogeny of these sequences was generated using RAxML [65,66] with 38 subtype reference *pol* sequences from the Los Alamos HIV Database (www.hiv.lanl.gov) used as an outgroup.

Table 1. Demographics of patients whose samples were analyzed.

		Subtype B (n = 8,483)
Age at Set-point (years) (mean, range):		35.4 (15–83)
Log10 Set-point Viral Load (mean, SD):		4.493±0.86
Sex	Female:	464 *(5.5%)*
	Male:	8019 *(94.5%)*
Risk Group	Homo/Bisexual:	7278 *(85.8%)*
	Heterosexual:	711 *(8.4%)*
	IDU:	239 *(2.8%)*
	Other/Unknown:	255 *(3.0%)*
Ethnicity	White:	6990 *(82.4%)*
	Black:	597 *(7.0%)*
	Asian:	221 *(2.6%)*
	Other/Unknown:	675 *(8.0%)*

doi:10.1371/journal.ppat.1004112.t001

Preliminary runs in ASReml were used to determine the fixed and random effects for the model. Sex, ethnicity, country of origin, age when the set-point viral load was taken, year of HIV diagnosis, and time from HIV diagnosis to the date when set-point viral load was taken, were all included in the final model (effect estimates given in Table S1). Set-point viral load was found to increase with age, but decrease with a more recent year of diagnosis and with a longer time period between HIV diagnosis and viral load testing. HIV-positive females and non-white individuals were found to have decreased set-point viral load measures compared to males and white individuals. The random effects were estimated to have a variance of 3.11×10^{-3} and 6.55×10^{-4} for year of HIV diagnosis and country of origin, respectively.

To confirm that our method performed as expected when tested on trees with known heritabilities we performed a simulation analysis similar to that of Alizon et al. [50]. We found the estimated heritability values to correspond well with the simulated values (see Text S1).

Bootstrapped phylogenetic trees were reconstructed in duplicate on the 8,483 sequences and both trees analyzed using ASReml independently. Using the comparison of the resulting log-likelihood values from running the model with and without the tree to estimate significance, both replicates produced highly significant ($p<0.0001$) heritability estimates of 5.8% (CI 2.9–8.7%) and 5.6% (CI 2.6–8.5%; Table 2).

As is typical for phylogenies based on population samples of HIV *pol* sequences, there is relatively little well-supported internal structure. In order to avoid possible bias in the heritability estimates, the analysis was repeated after splits with bootstrap-support values less than 90% were collapsed (Fig. S1), which removed 78% of internal nodes. Nevertheless, the heritability estimates remained significant in each case, with estimates of 5.1% (CI 2.4–7.8%) and 6.0% (CI 3.1–8.8%). However, when the entire tree was collapsed (excepting the split to the outgroup) leaving only branch-length information, the estimate was not significant, highlighting that detecting the heritability signal relies on at least some tree structure. One hundred bootstrapped phylogenies were analyzed to further examine the effect of uncertainty in the tree. Only four of the resulting heritability estimates failed to reach significance after Bonferroni correction (though their p-values were still <0.002), resulting in a mean heritability estimate of 5.5% (CI 2.6–8.5%) (data not shown).

In order to investigate how the viral genetic effect on set-point viral load varies across the phylogeny and through time, we constructed a time-resolved phylogeny using BEAST [67]. For reasons of computational tractability, this phylogeny had to be generated on a 652-sequence sub-sample of the dataset but produced a significant heritability estimate of 5.1% (upper CI 11.2%, $p<0.005$). We then used ASReml to estimate the phylogenetic effect of each node on viral load and mapped these estimates onto the time-resolved phylogeny, allowing the distribution of the effects across the tree and over time to be visualized. This showed some viral lineages to be clearly associated with substantial positive genetic effects on viral load, relative to the mean, and others to be associated with equally large negative effects (Fig. 1).

To investigate more formally the change in set-point viral load over time, we conducted an analysis in the R package MCMCglmm [68,69] in order to estimate the change in viral load due to selection on the virus and environmental effects using information from the temporal variation in sample dates (see Text S2). Analysis of the change due to selection on the virus and environmental effects revealed that this would have contributed a small but significant negative change in viral load of -0.05 $\log_{10}$ copies/mL/year (Fig. 2).

Discussion

Our analysis showed that viral genotype has a small but significant effect on set-point viral load in this population, with an estimated mean heritability of 5.7% (CI 2.8–8.6%). When the analysis was repeated after subsampling and using a different phylogenetic method, the heritability remained significant and did not differ significantly from the original estimate. As the star-like

Table 2. Estimates of viral genetic influence on set-point viral load in HIV subtype B in the UK.

Dataset	Method	*N*	Replicate	Viral Heritability (Conf. Interval)	
Full dataset	RAxML	8,483	1	5.8%	(2.9–8.7%)
			2	5.6%	(2.6–8.5%)
Nodes with bootstraps <90% collapsed	RAxML	8,483	1	5.1%	(2.4–7.8%)
			2	6.0%	(3.1–8.8%)
BEAST 652 Sub-Sample	BEAST	652	1	5.1%	(0–11.2%)
1,726 sequences with only 1 viral load removed	RAxML	6,757	1	7.8%	(4.3–11.3%)
			2	6.6%	3.4–9.9%)

doi:10.1371/journal.ppat.1004112.t002

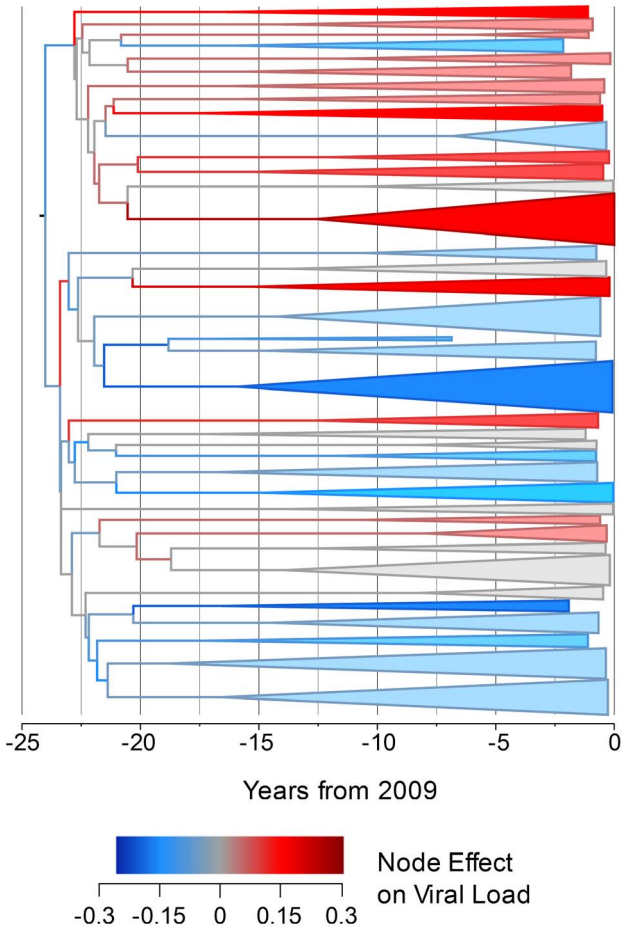

Figure 1. The estimated node effect plotted onto the phylogeny. The estimated phylogenetic effect of each node on $\log_{10}$ viral load plotted back onto the phylogeny from the 652-sample BEAST analysis. The axis shows the time in years from the most recent sequence, which was taken in 2009. Branches have been colored by the scale of the effect. Clusters of branches have been collapsed to improve readability, and are colored by the average tip effect within each cluster. As the number of bifurcations in the tree reduces at around 17.5 years before 2009, this used as the threshold for collapsing. Nodes that have a similar effect on viral load cluster together, as expected if some of the variation in viral load is heritable.
doi:10.1371/journal.ppat.1004112.g001

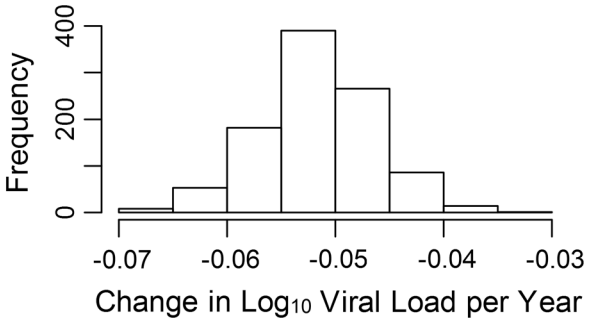

Figure 2. Change in viral load over time due to selection. The estimated $\log_{10}$ change in viral load per year due to selection and environmental effects (see also Fig. S2).
doi:10.1371/journal.ppat.1004112.g002

structure of HIV phylogenies can cause poor resolution of the internal nodes, resulting in low split support values, the impact of this effect was tested by collapsing weakly-supported nodes and analyzing one hundred bootstrapped phylogenies. This showed that the heritability estimates and their significance were not due to spurious or poorly-supported splits. Finally, a simulation analysis following the method of Alizon et al. [50] confirmed that our method of estimating heritability down a phylogeny performed as expected on a phylogeny where heritability is known (Text S1).

Analyzing smaller sampled datasets in BEAST allowed further investigation of the genetic effect on viral load. Plotting the estimated node effect on viral load back onto the phylogeny for the 652 sampled sequences illustrates the association of closely related sequences and similar genetic effects on viral loads in transmission chains that seem to have begun differentiating around the time subtype B arrived in the UK [70]. Finding viral lineages with both positive and negative genetic effects on viral load indicates that there is viral genetic variation that acts to both increase and decrease viral load relative to the mean.

Our estimates of the fixed effects influencing set-point viral load reflect previous reports identifying age [71,72] and sex [73–75] as significant, with older individuals and males having higher set-point viral loads. We also found ethnicity to have a significant effect on set-point viral load, finding a similar estimate for the effect of Black-African ethnicity to a previous paper looking specifically at this effect [76]. Although many previous studies on the influence of ethnicity on set-point viral load suggest there is no difference between ethnic groups or that non-white minorities have higher viral loads [77–79], differences in socio-economic status, risk-group, and access to care make the effect of ethnicity difficult to investigate [79]. Our finding that those with a longer time from HIV diagnosis to viral load testing had a slightly lower set-point viral load could reflect that individuals with lower viral load progress more slowly and therefore may be in general slower to access care, and also indicates that we are not classifying late-stage, rising viral loads as 'set-point,' which would result in the opposite effect. Finally, the fact that individuals with a more recent year of diagnosis also have a slightly lower set-point viral load could suggest that the proportion of individuals being diagnosed in late-stage infection is decreasing with time [80].

Previous studies investigating the heritability of viral load in HIV have estimated the genetic effect at between 11 to 60%, higher than the estimate of 5.7% obtained here. Three of the five studies were done on cohorts infected with subtypes other than B; one on subtype C [10] and two on mixed subtype populations [47,49], making comparisons difficult. Because virulence differs between subtypes [44–46], heritability estimates could be affected in studies where the cohort is infected with multiple subtypes, even when subtype is included as a variable in the model. Similarly, both the environmental and genetic variance that determines heritability can vary between populations, and may be particularly divergent between studies focusing on different demographic or risk groups (see [81] for further discussion). Considering this, some disparity in heritability estimates may not be unexpected.

Four of the previous studies used transmission pairs ($n = 28$ to $n = 115$) to estimate the heritability of viral load, and this could also influence the estimates obtained. As pointed out in one of these studies [10], the sero-discordant couples where transmission does occur may not accurately reflect the epidemic as a whole. As viral load is proportional to the probability of transmission, partners who transmit HIV and thus get included in the analysis had higher viral loads than average for the study [10]. Cohabiting or long-term sexual partners may also share confounded environmental factors such as diet and exposure to other

pathogens, which could affect health and thus viral load, and may even share HLA alleles, which increases HIV transmission risk and the between-partner correlation in viral load [82,83].

The only previous study that utilized a phylogeny-based approach also reported a heritability estimate considerably higher than the one obtained here. Alizon et al. [50] obtained a significant heritability estimate of around 50% when they used the most stringent criteria to define which samples would be taken as set-point viral load. Heritability estimates apply only to the population studied, so their estimate may be specific to this small (n = 134) population of MSM individuals with exceptionally stable viral load measures. Interestingly, when they relaxed their definition of set-point viral load, tripling the sample size, the heritability estimates shrank to around 11%. More generally, heritability estimates between studies where the sequences have different times to their respective MRCA are not readily compared. Studies with a more distant MRCA are likely to have higher heritability estimates as we expect the variance of the phenotype at the tips to increase with increasing time to the MRCA.

Given that viral genotype is influencing viral load, the question arises as to the source of this effect in the viral genome. The analysis has been performed on the *pol* gene, where both drug resistant and naturally occurring variation is known to affect replicative capacity [84]. It is also possible that this between-lineage variation (Fig. 1) could be a distal effect that may map to one or more other genes, such as *env* [85], that we are detecting through its linkage with variants in the *pol* gene. With increasing availability of full-genome datasets it may be possible to address that question directly in future.

The analysis performed here avoided issues associated with using multiple subtypes, transmission pairs, or restricted samples by including as many cases as possible. The aim was to minimize bias, but this clearly would be expected to introduce a substantial amount of noise and depends on the availability of large datasets. In fact twenty-fold more individuals were included than the largest previous dataset with a significant heritability estimate [50]. Nevertheless, this approach could allow some viral loads to be classified as set-point when they were actually taken during the acute stage, prior to the onset of AIDS, while on ART, or during a transient rise in viral load. The data cleaning methods utilized were able to exclude several cases that may have fallen into these categories, but this was difficult when there was only one pre-ART measure, as applied to approximately 20% of the dataset (1,726 cases). If many of the viral loads classified as set-point are not actually set-point measurements, this could affect the estimate of heritability obtained. However, when the dataset was re-run after removing these 1,726 cases, the heritability estimates remained significant with a mean value of 7.2% (CI 3.9–10.6%), showing that any errors made in classifying sequences with just one pre-ART viral load do not significantly affect the estimate.

We found no evidence that subtype B HIV is becoming more virulent in the UK. Indeed, the relatively small heritability of around 6% implies that host, environmental, and demographic effects play a much larger role in determining viral load than the virus genotype in this population and suggests that any change in viral load due to the viral genotype would be relatively small. The implications of a heritable viral load have been extensively explored, especially in the context of HIV adapting towards an 'optimal' viral load for transmission due to selection [2,81]. Our findings, however, imply that selection on the viral genetic component of viral load would have very limited influence on viral evolution.

The MCMCglmm analysis estimated a small but significant decrease over time of -0.05 $\log_{10}$ copies/mL/year in the mean value of the component of viral load determined by viral genotype (see Text S2). At this time the change due to selection on the virus cannot be disentangled from change to due environmental effects we have not controlled for, such as the background level of ART in the population, so we cannot assume all (or even any) of this change is due to selection on the viral genome. It should also be noted that though the viral genetic influence on viral load seems to be causing a decrease in viral load, this does not necessarily mean that overall viral load would be expected to decrease. As we estimate the viral genetic contribution to the variance in viral load to be only about 6%, changes in any of the many host and environmental factors influencing viral load could cause viral load to remain constant or even increase.

Previous cohort-based studies of viral load data have indeed estimated an increase in the phenotypic value. In an analysis based on 1,584 individuals with viral load data from the 22 CASCADE cohorts, Dorrucci et al. [36] estimated an increase in set-point of 0.044 $\log_{10}$ copies/mL/year, leading to an increase in set-point viral load of more than a log over 30 years. Herbeck et al. [37] performed a meta-analysis based on eight previous studies investigating change in viral load, which generated a more modest estimate of 0.013 $\log_{10}$ copies/mL/year and an overall increase of 0.39 $\log_{10}$ copies/mL in 30 years. These changes have led to suggestions that the virus may have evolved to become more virulent [36,37], but this was not directly analyzed and is clearly not the case in our study. However, a much larger fraction of the phenotypic value of viral load in our model is determined by the fixed effects including sex, age and time from diagnosis to first viral load which have certainly not remained constant over the course of the epidemic, so the two observations by no means necessarily conflict. The studies included by Herbeck et al. range from -0.013 to 0.056 $\log_{10}$ copies/mL/year in their estimates, with the largest study reporting a significant decline of -0.013 $\log_{10}$ copies/mL/year. This suggests that changes in viral load are difficult to quantify and may be quite population specific, with different environmental effects and selection pressures working in each.

Our findings indicate that the genotype of HIV subtype B in the UK has a small but significant effect on viral load, and suggest that the virulence of HIV has not increased. The use of this novel method in other situations where sequence data are available could allow estimation of heritability where it has not previously been possible.

Methods

8,700 initial subtype B sequences from the UK HIV RDB had viral load measures before starting ART available from UK CHIC. Sequences were aligned using the Stanford HIVdb Program [86], with manual checks for high levels of ambiguity and poor quality. To maintain both the representativeness of the HIV epidemic in the UK and as large a sample size as possible to improve power, a liberal definition of set-point viral load was chosen. If multiple pre-ART viral load measures were available for a patient, the first viral load was generally taken as the 'set-point' viral load. To exclude viral loads taken in AIDS, while the patient was on unreported ART, or while the patient was still in the acute phase of the disease, exclusion rules were applied. Unusually low or high viral load measures ($<$400 copies/mL or $>1\times10^6$ copies/

mL) were inspected for evidence of unreported ART use, acute infection, or onset of AIDS, and excluded if any of these were suggested. A full description of the rules used to discard records and the number of records removed can be found in Table S2. Set-point viral load values were $\log_{10}$ transformed before the analysis to make the distribution approximately normal. 80% of the patients included in our dataset had the viral load used as set-point taken within three years of HIV diagnosis. More information about the dates of HIV diagnosis and set-point viral load tests is available in Text S3.

Patients with accepted viral loads and at least 840 nucleotides of *pol* sequence (PR and partial RT coding regions) were analyzed. A total of 119 identical sequences from different patients were also removed, as they are likely to include multiple sequences from the same individual submitted under different identifiers. This left 8,483 subtype B sequences with matched viral loads. Sequences were stripped of codons in positions associated with drug-resistance mutations [87,88] before phylogenetic analysis. The analysis was repeated on sequences not stripped of resistance-associated codons, but no significant differences in heritability estimates were observed. To provide an unbiased root for the tree, 38 subtype reference *pol* sequences (subtypes A-K) from the Los Alamos HIV Database (www.hiv.lanl.gov) were used as an outgroup.

The large size of the dataset limited the methods available to create the phylogenies. RAxML [65,66] is an ML-based phylogenetic program designed to handle large alignments and produce accurate phylogenies by conducting a thorough topology search [89] and also performing bootstraps. We implemented RAxML on the Edinburgh Compute and Data Facility computer cluster. 100 bootstraps for each phylogeny were generated using the parallelized version of RAxML on 16 processors with a run time of 30 hrs. A comprehensive ML tree search was performed using the threaded version of RAxML on 12 processors for an average of 100 hrs, and the bootstrap-support values were then written onto the ML tree.

Pipeline

A new piece of software, TreeCollapseCL 4 (available at http://hiv.bio.ed.ac.uk), was developed to aid in preparing the phylogenies for further analysis and investigation of the data. Using TreeCollapseCL 4, each phylogeny was rooted and the average length of the tree was calculated from the tips to the MRCA of the UK sequences in the dataset (the second node from the root). Branch length to the root was not calculated because the distance from the root to the MRCA of the UK HIV RDB sequences can be severely affected by the choice of outgroup used.

The sampled viral sequences and all of the internal nodes of the phylogeny were incorporated into a genetic relatedness matrix from which the inverse was calculated using the R [68] package MCMCglmm [69].

The phylogenetic covariance of two individuals on a phylogeny was assumed to be proportional to the distance between their MRCA and the root [59]. Thus the covariance of an individual with itself is its distance from the root in units of substitutions per site per year. In phylogenetic comparative methods that use ultrametric trees, the distance between the tips and the root is often rescaled to one unit. Although the units are arbitrary, the variance explained by the phylogeny is directly interpretable as the variance explained in the sample of individuals used in the analysis. However, when trees are not ultrametric, as in this case, the root-to-tip distances vary. In this instance we standardized by the average distance from root to tip of 0.14 substitutions per site per year calculated earlier by TreeCollapseCL 4.

Preliminary runs were carried out on the dataset in ASReml in order to identify the fixed effects to include in the final model. Age at the sample date taken for set-point viral load, sex, ethnicity, time from HIV diagnosis to the set-point viral load sample date, and year of HIV diagnosis (as a continuous effect) were included in the preliminary models. All of the terms were found to be highly significant ($p<0.001$) and therefore all were included in the final model. Country of origin and year of HIV diagnosis (as a categorical effect) were also included as random effects along with the phylogeny. Year of HIV diagnosis was included as a continuous fixed effect to model the linear change in set-point viral load and as a categorical random effect to account for any random deviations around this trend from year to year. As country of origin has many discrete levels, it was included as a random effect in order to estimate the variance of their effects.

The significance of the effect of the phylogeny in explaining the variance was assessed by first running the model without the phylogeny as a 'null' model and then including the phylogeny. A log-likelihood ratio test with one degree of freedom was then used to test whether the model with the phylogeny was significantly better at explaining the variation in viral load than the null model.

As ASReml assumes all pedigree information provided is correct, analyses were repeated using TreeCollapseCL 4 to collapse splits with bootstrap-support values less than 90% down to polytomies. To further evaluate how uncertainty in the tree could affect our heritability estimates, one hundred bootstrapped trees were generated in RAxML and analyzed. Because of the close phylogenetic relationship between the subtype B and D in the *pol* region bootstrapped subtype D sequences can sometimes cluster within the B clade, making it necessary to remove the subtype D Los Alamos sequences from the phylogeny in order to root all 100 trees by the same outgroup.

Each analysis was performed in duplicate, with the sequences being run through RAxML and the analysis pipeline twice. The significance threshold used was adjusted using a Bonferroni correction for the number of replicates.

In order to investigate whether set-point viral load has changed over time, we estimated the amount of change in viral load due to selection. This can be estimated using Markov chain Monte Carlo methods to calculate the total contribution of between-lineage and within-host selection, though we cannot distinguish all change due to within-host selection from environmental factors (see Text S2 for more detail).

In addition, to further investigate the phylogenetic effects on viral load, a time-scaled phylogeny was produced using BEAST, a Bayesian phylogenetic program [67]. Because the complexity of the analysis performed by BEAST limits the number of samples which can be run in a reasonable time-frame, a sub-sample of the main dataset was used.

After collapsing nodes with bootstrap support values less than 90%, 965 sequences remained in un-collapsed clusters of fifteen or more sequences. A random subsample of 652 of these sequences was taken for analysis in BEAST. BEAST was run with a relaxed log-normal clock and a constant population size for 100,000,000 steps, sampling every 10,000 steps. All runs were performed in duplicate, and after 10% burn-in was removed the resulting files were combined using an in-house script. A summary tree was then generated using the BEAST program TreeAnnotator, and run in ASReml to obtain heritability estimates.

Finally, in order to validate the heritability estimates produced by our pipeline, we followed the method of Alizon et al. [50] (see Text S1) to perform a simulation analysis where viral loads were simulated down trees under known heritabilities.

Accession numbers

As submission of the entire UK HIV Drug Resistance Database online would risk breaching patient confidentiality by allowing transmission networks to be identified, following Kouyos et al. (J Infect Dis, 2010) and Leigh Brown et al. (J Infect Dis, 2011) a random sample of 10% of the database has been submitted to GenBank under accession numbers JN100661-JN101948.

Ethics statement

This work was performed on data generated in the course of routine clinical care which was anonymized and delinked before analysis. Ethical approval for this work was given by the London Multicentre Research Ethics Committee (MREC/01/2/10; 5 April 2001).

Members of the UK HIV Drug Resistance Database Steering Committee

Celia Aitken (Gartnavel General Hospital, Glasgow); David Asboe, Anton Pozniak (Chelsea & Westminster Hospital, London); Patricia Cane (Public Health England, Porton Down); Hannah Castro, David Dunn*, Esther Fearnhill, Kholoud Porter (MRC Clinical Trials Unit at UCL, London); David Chadwick (South Tees Hospitals NHS Trust, Middlesbrough); Duncan Churchill (Brighton and Sussex University Hospitals NHS Trust); Duncan Clark (St Bartholomew's and The London NHS Trust); Simon Collins (HIV i-Base, London); Valerie Delpech (Centre for Infections, Public Health England); Samuel Douthwaite (Guy's and St. Thomas' NHS Foundation Trust, London); Anna Maria Geretti (Institute of Infection and Global Health, University of Liverpool); Antony Hale (Leeds Teaching Hospitals NHS Trust); Stéphane Hué (University College London); Steve Kaye (Imperial College, London); Paul Kellam (Wellcome Trust Sanger Institute & University College London Medical School); Linda Lazarus (Expert Advisory Group on AIDS Secretariat, Public Health England); Andrew Leigh-Brown (University of Edinburgh); Tamyo Mbisa (Virus Reference Department, Public Health England); Nicola Mackie (Imperial NHS Trust, London); Chloe Orkin (St. Bartholomew's Hospital, London); Eleni Nastouli, Deenan Pillay*, Andrew Phillips, Caroline Sabin (University College London Medical School. London); Erasmus Smit (Public Health England, Birmingham Heartlands Hospital); Kate Templeton (Royal Infirmary of Edinburgh); Peter Tilston (Manchester Royal Infirmary); Daniel Webster (Royal Free NHS Trust, London); Ian Williams (Mortimer Market Centre, London); Hongyi Zhang (Addenbrooke's Hospital, Cambridge); Mark Zuckerman (King's College Hospital, London).

*Co-PI

Members of the UK CHIC Steering Committee

Jonathan Ainsworth, North Middlesex University Hospital NHS Trust, London; Sris Allan, Coventry & Warwickshire NHS Trust; Jane Anderson, Homerton University Hospital NHS Trust, London; Abdel Babiker, MRC Clinical Trials Unit, London; David Chadwick, South Tees Hospitals NHS Foundation Trust, Middlesbrough; Valerie Delpech, Health Protection Agency Centre for Infections (HPA CfI), London; David Dunn, MRC Clinical Trials Unit, London; Martin Fisher, Brighton and Sussex University Hospitals NHS Trust, Brighton; Brian Gazzard (Chair), Chelsea & Westminster Hospital NHS Foundation Trust, London; Richard Gilson, Mortimer Market Centre, University College London Medical School; Mark Gompels, North Bristol NHS Trust, Bristol; Phillip Hay, St George's Healthcare NHS Trust, London; Teresa Hill, University College London Medical School; Margaret Johnson, Royal Free Hampstead NHS Trust, London; Stephen Kegg, South London Healthcare NHS Trust, London; Clifford Leen, The Lothian University Hospitals NHS Trust, Edinburgh; Fabiola Martin, York Teaching Hospital NHS Foundation Trust; Mark Nelson, Chelsea and Westminster Hospital NHS Foundation Trust, London; Chloe Orkin, Barts and The London NHS Trust, London; Adrian Palfreeman, University Hospitals of Leicester NHS Trust; Andrew Phillips, University College London Medical School; Deenan Pillay, University College London; Jillian Pritchard, Ashford & St. Peter's Hospitals NHS Foundation Trust; Frank Post, King's College Hospital NHS Foundation Trust, London; Caroline Sabin, University College London Medical School; Memory Sachikonye, UK Community Advisory; Board (UK-CAB); Achim Schwenk, North Middlesex University Hospital NHS Trust, London; Anjum Tariq, The Royal Wolverhampton NHS Trust; John Walsh, Imperial College Healthcare NHS Trust, London.

Supporting Information

Figure S1 The effect of collapsing poorly-supported nodes. A sub-section of the full RAxML tree is shown before (A) and after (B) collapsing nodes with bootstrap support less than 90% down to polytomies. Branch length from root to tip nodes is preserved after collapsing.
(TIF)

Figure S2 The $\log_{10}$ change in viral load per year due to selection. The estimated $\log_{10}$ change in viral load per year due to between-lineage selection (shaded) and within-host selection and environmental effects (unshaded). Though the change due to between-lineage selection was not significantly different from what could be expected through drift, the change due to within-host selection and environmental effects was significant.
(TIF)

Table S1 Mean fixed effect estimates.
(PDF)

Table S2 Details of sequences removed during data cleaning.
(PDF)

Text S1 Supplementary methods detailing the simulation performed to verify the heritability estimate obtained.
(PDF)

Text S2 Within-host and between-lineage selection analysis and simulations.
(PDF)

Text S3 Additional information about time of diagnosis and viral load test date.
(PDF)

Acknowledgments

We would like to thank the reviewers and editors for their helpful suggestions, Dr. S. Reece for originally raising this question with ALB, and HH for contributing an atmosphere that encouraged this collaboration. We would also like to thank Albert Phillimore for useful discussions and Samuel Alizon for providing us with his simulated phylogenies. This work has made use of the resources provided by the Edinburgh Compute and Data Facility (ECDF) (?http://www.ecdf.ed.ac.uk/).

Author Contributions

Conceived and designed the experiments: EH JDH AP AJLB. Performed the experiments: EF SO. Analyzed the data: EH JDH EF AJLB. Contributed reagents/materials/analysis tools: JDH SO DD DP. Wrote the paper: EH JDH AJLB.

References

1. Mellors JW, Rinaldo CR, Gupta P, White RM, Todd JA, et al. (1996) Prognosis in HIV-1 Infection Predicted by the Quantity of Virus in Plasma. Science 272: 1167–1170. doi:10.1126/science.272.5265.1167.
2. Fraser C, Hollingsworth TD, Chapman R, de Wolf F, Hanage WP (2007) Variation in HIV-1 set-point viral load: Epidemiological analysis and an evolutionary hypothesis. Proceedings of the National Academy of Sciences 104: 17441–17446. doi:10.1073/pnas.0708559104.
3. Langford SE, Ananworanich J, Cooper DA (2007) Predictors of disease progression in HIV infection: a review. AIDS Res Ther 4: 11. doi:10.1186/1742-6405-4-11.
4. Quinn TC, Wawer MJ, Sewankambo N, Serwadda D, Li C, et al. (2000) Viral Load and Heterosexual Transmission of Human Immunodeficiency Virus Type 1. N Engl J Med 342: 921–929. doi:10.1056/NEJM200003303421303.
5. Fideli US, Allen SA, Musonda R, Trask S, Hahn BH, et al. (2001) Virologic and immunologic determinants of heterosexual transmission of human immunodeficiency virus type 1 in Africa. AIDS Res Hum Retroviruses 17: 901–910. doi:10.1089/088922201750290023.
6. Wawer MJ, Gray RH, Sewankambo NK, Serwadda D, Li X, et al. (2005) Rates of HIV-1 Transmission Per Coital Act, by Stage of HIV-1 Infection, in Rakai, Uganda. J Infect Dis 191: 1403–1409. doi:10.1086/429411.
7. Steel CM, Beatson D, Cuthbert RJG, Morrison H, Ludlam CA, et al. (1988) HLA Haplotype A1 B8 DR3 as a Risk Factor for HIV-Related Disease. The Lancet 331: 1185–1188. doi:10.1016/S0140-6736(88)92009-0.
8. Kaslow RA, vanRaden M, Friedman H, Duquesnoy R, Marrari M, et al. (1990) A1, Cw7, B8, DR3 HLA antigen combination associated with rapid decline of T-helper lymphocytes in HIV-1 infection: A report from the Multicenter AIDS Cohort Study. The Lancet 335: 927–930. doi:10.1016/0140-6736(90)90995-H.
9. O'Brien SJ, Nelson GW (2004) Human genes that limit AIDS. Nat Genet 36: 565–574. doi:10.1038/ng1369.
10. Tang J, Tang S, Lobashevsky E, Zulu I, Aldrovandi G, et al. (2004) HLA allele sharing and HIV type 1 viremia in seroconverting Zambians with known transmitting partners. AIDS Res Hum Retroviruses 20: 19–25. doi:10.1089/088922204322749468.
11. Fellay J, Ge D, Shianna KV, Colombo S, Ledergerber B, et al. (2009) Common Genetic Variation and the Control of HIV-1 in Humans. PLoS Genet 5: e1000791. doi:10.1371/journal.pgen.1000791.
12. Salgado M, Brennan T, O'Connell K, Bailey J, Ray S, et al. (2010) Evolution of the HIV-1 nef gene in HLA-B*57 Positive Elite Suppressors. Retrovirology 7: 94. doi:10.1186/1742-4690-7-94.
13. Huang Y, Paxton WA, Wolinsky SM, Neumann AU, Zhang L, et al. (1996) The role of a mutant CCR5 allele in HIV-1 transmission and disease progression. Nat Med 2: 1240–1243. doi:10.1038/nm1196-1240.
14. Smith MW, Dean M, Carrington M, Winkler C, Huttley GA, et al. (1997) Contrasting Genetic Influence of CCR2 and CCR5 Variants on HIV-1 Infection and Disease Progression. Science 277: 959–965. doi:10.1126/science.277.5328.959.
15. Pido-Lopez J, Whittall T, Wang Y, Bergmeier LA, Babaahmady K, et al. (2007) Stimulation of Cell Surface CCR5 and CD40 Molecules by Their Ligands or by HSP70 Up-Regulates APOBEC3G Expression in CD4+ T Cells and Dendritic Cells. J Immunol 178: 1671–1679.
16. Åsjö B, Albert J, Karlsson A, Morfeldt-Månson L, Biberfeld G, et al. (1986) Replicative capacity of human immunodeficiency virus from patients with varying severity of HIV infection. The Lancet 328: 660–662. doi:10.1016/S0140-6736(86)90169-8.
17. Fenyo EM, Morfeldt Manson L, Chiodi F, Lind B, von Gegerfelt A, et al. (1988) Distinct replicative and cytopathic characteristics of human immunodeficiency virus isolates. JVirol 62: 4414–4419.
18. Fiore JR, Calabro ML, Angarano G, De Rossi A, Fico C, et al. (1990) HIV-1 variability and progression to AIDS: a longitudinal study. J Med Virol 32: 252–256.
19. Hutchinson CM, Wilson C, Reichart CA, Marsiglia VC, Zenilman JM, et al. (1991) CD4 Lymphocyte Concentrations in Patients With Newly Identified HIV Infection Attending STD Clinics: Potential Impact on Publicly Funded Health Care Resources. JAMA 266: 253–256. doi:10.1001/jama.1991.03470020079036.
20. Weiss PJ, Brodine SK, Goforth RR, Kennedy CA, Wallace MR, et al. (1992) Initial Low CD4 Lymphocyte Counts in Recent Human Immunodeficiency Virus Infection and Lack of Association with Identified Coinfections. The Journal of Infectious Diseases 166: 1149–1153.
21. Gorham ED, Garland FC, Mayers DL, Goforth RR, Brodine SK, et al. (1993) CD4 Lymphocyte Counts Within 24 Months of Human Immunodeficiency Virus Seroconversion: Findings in the US Navy and Marine Corps. Arch Intern Med 153: 869–876. doi:10.1001/archinte.1993.00410070055008.
22. Holmberg SD, Conley LJ, Luby SP, Cohn S, Wong LC, et al. (1995) Recent Infection with Human Immunodeficiency Virus and Possible Rapid Loss of CD4 T Lymphocytes. JAIDS Journal of Acquired Immune Deficiency Syndromes 9: 291–296.
23. Veugelers PJ, Page KA, Tindall B, Schechter MT, Moss AR, et al. (1994) Determinants of HIV Disease Progression among Homosexual Men Registered in the Tricontinental Seroconverter Study. American Journal of Epidemiology 140: 747–758.
24. O'Brien TR, Hoover DR, Rosenberg PS, Chen B, Detels R, et al. (1995) Evalution of Secular Trends in CD4+ Lymphocyte Loss among Human Immunodeficiency Virus Type 1 (HIV-1)-infected Men with Known Dates of Seroconversion. American Journal of Epidemiology 142: 636–642.
25. Galai N, Lepri AC, Vlahov D, Pezzotti P, Sinicco A, et al. (1996) Temporal Trends of Initial CD4 Cell Counts Following Human Immunodeficiency Virus Seroconversion in Italy, 1985–1992. American Journal of Epidemiology 143: 278–282.
26. Keet IPM, Veugelers PJ, Koot M, de Weerd MH, Roos MTL, et al. (1996) Temporal trends of the natural history of HIV-1 infection following seroconversion between 1984 and 1993. AIDS 10: 1601–1602.
27. Carré N, Prins M, Meyer L, Brettle RP, Robertson JR, et al. (1997) Has the rate of progression to AIDS changed in recent years? AIDS 11: 1611–1618.
28. Sinicco A, Fora R, Raiteri R, Sciandra M, Bechis G, et al. (1997) Is the clinical course of HIV-1 changing? Cohort study. BMJ 314: 1232–1237.
29. Vanhems P, Lambert J, Guerra M, Hirschel B, Allard R (1999) Association between the rate of CD4+ T cell decrease and the year of human immunodeficiency virus (HIV) type 1 seroconversion among persons enrolled in the Swiss HIV cohort study. J Infect Dis 180: 1803–1808. doi:10.1086/315110.
30. Concerted Action on SeroConversion to AIDS and Death in Europe (2000) Time from HIV-1 seroconversion to AIDS and death before widespread use of highly-active antiretroviral therapy: a collaborative re-analysis. The Lancet 355: 1131–1137. doi:doi: DOI: 10.1016/S0140-6736(00)02061-4.
31. CASCADE Collaboration (2003) Differences in CD4 cell counts at seroconversion and decline among 5739 HIV-1-infected individuals with well-estimated dates of seroconversion. J Acquir Immune Defic Syndr 34: 76–83.
32. Dorrucci M, Phillips AN, Longo B, Rezza G, The Italian Seroconversion Study(2005) Changes over time in post-seroconversion CD4 cell counts in the Italian HIV-Seroconversion Study: 1985–2002. AIDS 19: 331–335.
33. Müller V, Ledergerber B, Perrin L, Klimkait T, Furrer H, et al. (2006) Stable virulence levels in the HIV epidemic of Switzerland over two decades. AIDS 20: 889–894.
34. Crum-Cianflone N, Eberly L, Zhang Y, Ganesan A, Weintrob A, et al. (2009) Is HIV Becoming More Virulent? Initial CD4 Cell Counts among HIV Seroconverters During the Course of the HIV Epidemic: 1985–2007. Clin Infect Dis 48: 1285–1292. doi:10.1086/597777.
35. Müller V, Maggiolo F, Suter F, Ladisa N, De Luca A, et al. (2009) Increasing Clinical Virulence in Two Decades of the Italian HIV Epidemic. PLoS Pathog 5: e1000454. doi:10.1371/journal.ppat.1000454.
36. Dorrucci M, Rezza G, Porter K, Phillips A (2007) Temporal Trends in Postseroconversion CD4 Cell Count and HIV Load: The Concerted Action on Seroconversion to AIDS and Death in Europe Collaboration, 1985–2002. The Journal of Infectious Diseases 195: 525–534. doi:10.1086/510911.
37. Herbeck JT, Müller V, Maust BS, Ledergerber B, Torti C, et al. (2012) Is the virulence of HIV changing? A meta-analysis of trends in prognostic markers of HIV disease progression and transmission. AIDS 26: 193–205. doi:10.1097/QAD.0b013e32834db418.
38. Martinez-Picado J, Martínez MA (2008) HIV-1 reverse transcriptase inhibitor resistance mutations and fitness: A view from the clinic and ex vivo. Virus Research 134: 104–123. doi:10.1016/j.virusres.2007.12.021.
39. Fenner F, Chapple PJ (1965) Evolutionary changes in myxoma virus in Britain: An examination of 222 in naturally occurring strains obtained from 80 counties during the period October–November 1962. Journal of Hygiene 63: 175–185. doi:10.1017/S0022172400045083.
40. Anderson RM, May RM (1979) Population biology of infectious diseases: Part I. Nature 280: 361–367. doi:10.1038/280361a0.
41. May RM, Anderson RM (1983) Epidemiology and Genetics in the Coevolution of Parasites and Hosts. Proceedings of the Royal Society of London Series B, Biological Sciences 219: 281–313.
42. Korber B, Muldoon M, Theiler J, Gao F, Gupta R, et al. (2000) Timing the ancestor of the HIV-1 pandemic strains. Science 288: 1789–1796.
43. Worobey M, Gemmel M, Teuwen DE, Haselkorn T, Kunstman K, et al. (2008) Direct evidence of extensive diversity of HIV-1 in Kinshasa by 1960. Nature 455: 661–664. doi:10.1038/nature07390.
44. Kanki PJ, Hamel DJ, Sankalé J, Hsieh C, Thior I, et al. (1999) Human Immunodeficiency Virus Type 1 Subtypes Differ in Disease Progression. The Journal of Infectious Diseases 179: 68–73. doi:10.1086/314557.
45. Kaleebu P, Ross A, Morgan D, Yirrell D, Oram J, et al. (2001) Relationship between HIV-1 Env subtypes A and D and disease progression in a rural Ugandan cohort. AIDS 15: 293–299.
46. Kiwanuka N, Laeyendecker O, Robb M, Kigozi G, Arroyo M, et al. (2008) Effect of human immunodeficiency virus Type 1 (HIV-1) subtype on disease progression in persons from Rakai, Uganda, with incident HIV-1 infection. J Infect Dis 197: 707–713. doi:10.1086/527416.
47. Van der Kuyl AC, Jurriaans S, Pollakis G, Bakker M, Cornelissen M (2010) HIV RNA levels in transmission sources only weakly predict plasma viral load in recipients. AIDS 24: 1607–1608. doi:10.1097/QAD.0b013e32833b318f.
48. Hecht FM, Hartogensis W, Bragg L, Bacchetti P, Atchison R, et al. (2010) HIV RNA level in early infection is predicted by viral load in the transmission source. AIDS 24: 941–945. doi:10.1097/QAD.0b013e328337b12e.

49. Hollingsworth TD, Laeyendecker O, Shirreff G, Donnelly CA, Serwadda D, et al. (2010) HIV-1 Transmitting Couples Have Similar Viral Load Set-Points in Rakai, Uganda. PLoS Pathog 6: e1000876. doi:10.1371/journal.ppat.1000876.
50. Alizon S, von Wyl V, Stadler T, Kouyos RD, Yerly S, et al. (2010) Phylogenetic Approach Reveals That Virus Genotype Largely Determines HIV Set-Point Viral Load. PLoS Pathog 6: e1001123. doi:10.1371/journal.ppat.1001123.
51. Lynch M (1991) Methods for the Analysis of Comparative Data in Evolutionary Biology. Evolution 45: 1065–1080.
52. Pagel M (1999) Inferring the historical patterns of biological evolution. Nature 401: 877–884. doi:10.1038/44766.
53. Housworth EA, Martins EP, Lynch M (2004) The phylogenetic mixed model. Am Nat 163: 84–96. doi:10.1086/380570.
54. Patterson HD, Thompson R (1971) Recovery of Inter-Block Information When Block Sizes Are Unequal. Biometrika 58: 545–554. doi:10.1093/biomet/58.3.545.
55. Thompson R, Brotherstone S, White IMS, Thompson R, Brotherstone S, et al. (2005) Estimation of Quantitative Genetic Parameters. Phil Trans R Soc B 360: 1469–1477. doi:10.1098/rstb.2005.1676.
56. Hadfield JD, Nakagawa S (2010) General quantitative genetic methods for comparative biology: phylogenies, taxonomies and multi-trait models for continuous and categorical characters. Journal of Evolutionary Biology 23: 494–508. doi:10.1111/j.1420-9101.2009.01915.x.
57. Gilmour AR, Gogel BJ, Cullis BR, Thompson R (2009) ASReml User Guide Release 3.0. Available: www.vsni.co.uk.
58. Henderson CR (1976) A Simple Method for Computing the Inverse of a Numerator Relationship Matrix Used in Prediction of Breeding Values. Biometrics 32: 69–83. doi:10.2307/2529339.
59. Felsenstein J (1985) Phylogenies and the Comparative Method. AmNat 125: 1–15.
60. Hansen TF, Martins EP (1996) Translating Between Microevolutionary Process and Macroevolutionary Patterns: The Correlation Structure of Interspecific Data. Evolution 50: 1404–1417. doi:10.2307/2410878.
61. Davis TA (2006) Direct Methods for Sparse Linear Systems. SIAM. 229 p.
62. Freckleton RP (2012) Fast likelihood calculations for comparative analyses. Methods in Ecology and Evolution 3: 940–947. doi:10.1111/j.2041-210X.2012.00220.x.
63. Leigh Brown AJ, Lycett SJ, Weinert L, Hughes GJ, Fearnhill E, et al. (2011) Transmission Network Parameters Estimated From HIV Sequences for a Nationwide Epidemic. J Infect Dis 204: 1463–1469. doi:10.1093/infdis/jir550.
64. The UK Collaborative HIV Cohort Steering Committee (2004) The creation of a large UK-based multicentre cohort of HIV-infected individuals: The UK Collaborative HIV Cohort (UK CHIC) Study. HIV Medicine 5: 115–124. doi:10.1111/j.1468-1293.2004.00197.x.
65. Stamatakis A (2006) RAxML-VI-HPC: Maximum Likelihood-Based Phylogenetic Analyses with Thousands of Taxa and Mixed Models. Bioinformatics 22: 2688–2690. doi:10.1093/bioinformatics/btl446.
66. Stamatakis A, Blagojevic F, Nikolopoulos D, Antonopoulos C (2007) Exploring New Search Algorithms and Hardware for Phylogenetics: RAxML Meets the IBM Cell. The Journal of VLSI Signal Processing 48: 271–286. doi:10.1007/s11265-007-0067-4.
67. Drummond AJ, Suchard MA, Xie D, Rambaut A (2012) Bayesian Phylogenetics with BEAUti and the BEAST 1.7. Mol Biol Evol 29: 1969–1973. doi:10.1093/molbev/mss075.
68. R Development Core Team (2011) R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Available: http://www.R-project.org.
69. Hadfield JD (2010) MCMC Methods for Multi-Response Generalized Linear Mixed Models: The MCMCglmm R Package. Journal of Statistical Software 33: 1–22.
70. Hué S, Pillay D, Clewley JP, Pybus OG (2005) Genetic analysis reveals the complex structure of HIV-1 transmission within defined risk groups. Proc Natl Acad Sci USA 102: 4425–4429.
71. O'Brien TR, Blattner WA, Waters D, Eyster M, Hilgartner MW, et al. (1996) Serum HIV-1 RNA Levels and Time to Development of AIDS in the Multicenter Hemophilia Cohort Study. JAMA 276: 105–110. doi:10.1001/jama.1996.03540020027025.
72. Nogueras M, Navarro G, Antón E, Sala M, Cervantes M, et al. (2006) Epidemiological and clinical features, response to HAART, and survival in HIV-infected patients diagnosed at the age of 50 or more. BMC Infectious Diseases 6: 159. doi:10.1186/1471-2334-6-159.
73. Farzadegan H, Hoover DR, Astemborski J, Lyles CM, Margolick JB, et al. (1998) Sex differences in HIV-1 viral load and progression to AIDS. The Lancet 352: 1510–1514. doi:10.1016/S0140-6736(98)02372-1.
74. Sterling TR, Lyles CM, Vlahov D, Astemborski J, Margolick JB, et al. (1999) Sex Differences in Longitudinal Human Immunodeficiency Virus Type 1 RNA Levels among Seroconverters. J Infect Dis 180: 666–672. doi:10.1086/314967.
75. Gandhi M, Bacchetti P, Miotti P, Quinn TC, Veronese F, et al. (2002) Does Patient Sex Affect Human Immunodeficiency Virus Levels? Clin Infect Dis 35: 313–322. doi:10.1086/341249.
76. Müller V, von Wyl V, Yerly S, Böni J, Klimkait T, et al. (2009) African descent is associated with slower CD4 cell count decline in treatment-naive patients of the Swiss HIV Cohort Study. AIDS 23: 1269–1276. doi:10.1097/QAD.0b013e32832d4096.
77. Brown AE, Malone JD, Zhou SYJ, Lane JR, Hawkes CA (1997) Human Immunodeficiency Virus RNA Levels in US Adults: A Comparison Based upon Race and Ethnicity. J Infect Dis 176: 794–797. doi:10.1086/517304.
78. Swindells S, Cobos DG, Lee N, Lien EA, Fitzgerald AP, et al. (2002) Racial/ethnic differences in CD4 T cell count and viral load at presentation for medical care and in follow-up after HIV-1 infection. AIDS 16: 1832–1834.
79. Boyd A, Murad S, O'Shea S, De Ruiter A, Watson C, et al. (2005) Ethnic differences in stage of presentation of adults newly diagnosed with HIV-1 infection in south London. HIV Medicine 6: 59–65. doi:10.1111/j.1468-1293.2005.00267.x.
80. Health Protection Agency (2011) HIV in the United Kingdom: 2011 Report. London: Health Protection Services, Colindale. Available: http://www.hpa.org.uk/webc/HPAwebFile/HPAweb_C/1317131685847.
81. Müller V, Fraser C, Herbeck JT (2011) A Strong Case for Viral Genetic Factors in HIV Virulence. Viruses 3: 204–216. doi:10.3390/v3030204.
82. Lockett SF, Robertson JR, Brettle RP, Yap PL, Middleton D, et al. (2001) Mismatched Human Leukocyte Antigen Alleles Protect Against Heterosexual HIV Transmission. JAIDS Journal of Acquired Immune Deficiency Syndromes 27: 277–280.
83. Dorak MT, Tang J, Penman-Aguilar A, Westfall AO, Zulu I, et al. (2004) Transmission of HIV-1 and HLA-B allele-sharing within serodiscordant heterosexual Zambian couples. The Lancet 363: 2137–2139. doi:10.1016/S0140-6736(04)16505-7.
84. Hinkley T, Martins J, Chappey C, Haddad M, Stawiski E, et al. (2011) A systems analysis of mutational effects in HIV-1 protease and reverse transcriptase. Nat Genet 43: 487–489. doi:10.1038/ng.795.
85. Ariën KK, Troyer RM, Gali Y, Colebunders RL, Arts EJ, et al. (2005) Replicative fitness of historical and recent HIV-1 isolates suggests HIV-1 attenuation over time. AIDS 19: 1555–1564.
86. Liu TF, Shafer RW (2006) Web Resources for HIV Type 1 Genotypic-Resistance Test Interpretation. Clin Infect Dis 42: 1608–1618. doi:10.1086/503914.
87. Rhee S-Y, Gonzales MJ, Kantor R, Betts BJ, Ravela J, et al. (2003) Human immunodeficiency virus reverse transcriptase and protease sequence database. Nucleic Acids Research 31: 298–303.
88. Shafer RW (2006) Rationale and Uses of a Public HIV Drug-Resistance Database. The Journal of Infectious Diseases 194: S51–S58. doi:10.1086/505356.
89. Price MN, Dehal PS, Arkin AP (2010) FastTree 2 – Approximately Maximum-Likelihood Trees for Large Alignments. PLoS ONE 5: e9490. doi:10.1371/journal.pone.0009490.