How effectively can HIV phylogenies be used to measure heritability?

George Shirreff*[1,2], Samuel Alizon[3], Anne Cori[1], Huldrych F. Günthard[4], Oliver Laeyendecker[5,6], Ard van Sighem[7], Daniela Bezemer[7] and Christophe Fraser[1]

[1]Medical Research Council Centre for Outbreak Analysis and Modelling, Department of Infectious Disease Epidemiology, Imperial College, London, UK; [2]Institute for Integrative Biology, ETH Zürich, Zürich, Switzerland; [3]Lab MIVEGEC UMR CNRS 5290, IRD 224, UM1, UM2, Montpellier, France; [4]Division of Infectious Diseases and Hospital Epidemiology, University Hospital of Zürich, University of Zürich, Zürich, Switzerland; [5]National Institute of Allergy and Infectious Diseases, National Institutes of Health, Baltimore, MD, USA; [6]Department of Medicine, Johns Hopkins University School of Medicine, Baltimore, MD, USA; and [7]Stichting HIV Monitoring, Amsterdam, The Netherlands
*Correspondence address. Institute for Integrative Biology, ETH Zürich, Universitätstrasse 16, 8092 Zürich, Switzerland. Tel/Fax: +41(0)446320659; E-mail:george.shirreff@env.ethz.ch
Received 20 May 2013; revised version accepted 10 September 2013

ABSTRACT

Background and objectives: The severity of HIV-1 infection, measured by set-point viral load (SPVL), is highly variable between individuals. Its heritability between infections quantifies the control the pathogen genotype has over disease severity. Heritability estimates vary widely between studies, but differences in methods make comparison difficult. Phylogenetic comparative analysis offers measures of phylogenetic signal, but it is unclear how to interpret them in terms of the fraction of variance in SPVL controlled by the virus genotype.

Methodology: We present computational methods which link statistics summarizing phylogenetic signal to heritability, h^2 in order to test for and quantify it. We re-analyse data from Switzerland and Uganda, and apply it to new data from the Netherlands. We systematically compare established and new (e.g. phylogenetic pairs, PP) phylogenetic signal statistics.

Results: Heritability estimates varied by method and dataset. Several methods were consistently able to detect simulated heritability above $h^2 \approx 0.4$, but none below. Pagel's λ was the most robust and sensitive. The PP method found no heritability in the Netherlands data, whereas Pagel's λ found significant heritability only in a narrow subdivision ($P=0.038$). Heritability was estimated at $h^2=0.52$ (95% confidence interval 0.00–0.63).

Conclusions and implications: This standardized measure, h^2, allows comparability of heritability between cohorts. We confirm high heritability in Swiss data, but neither in Ugandan data nor in the Netherlands, where it is barely significant or undetectable. Existing phylogenetic methods are ill-suited for detecting heritability below $h^2 \approx 0.4$, which may nonetheless be biologically important.

KEYWORDS: HIV-1; heritability; phylogenetic comparative analysis; set-point viral load; phylogenetic signal; virulence

BACKGROUND AND OBJECTIVES

HIV has a high mutation rate [1, 2] and daily turnover [3, 4] and therefore adapts rapidly under local selective pressure from the immune system [5, 6, 7] or antiretroviral drugs [8, 9]. Increasingly there is interest in the transmission of escape [10, 11] or drug resistance mutations [12], which may enable viral adaptation to the host population. The rate at which a trait evolves in response to natural selection is determined by its heritability.

Recent work has suggested that virulence may also evolve at the population level [13] by natural selection towards a level optimal for transmission [14, 15]. Variation in virulence has a large impact on mortality and morbidity, so its evolutionary potential may present challenges or provide opportunities in public health interventions [16]. For example, vaccines which reduce growth rate or toxicity are predicted to reduce the costs of virulence, to the vaccinated host and the pathogen. This potentially raises the optimal virulence, resulting in poorer outcomes for the unvaccinated individuals [17].

Virulence in HIV is well approximated by set-point viral load (SPVL), which refers to the density of virions in the blood during asymptomatic infection. SPVL is an early prognostic indicator for AIDS, as it varies by orders of magnitude between individuals, with high values having faster CD4 cell decline, progressing more rapidly to AIDS and death [18, 19, 20]. However, it is relatively stable within the individual [21] meaning that it can be measured at a wide range of time points in an individuals' infection [22].

Many host factors influence SPVL (Human Leukocyte Antigen (HLA) type (reviewed in [23])), sex [24], ethnicity [25], age [26], co-infections [27, 28]). Specific human genetic markers have been identified to which 13% of SPVL variation can be attributed, with a further 9% explained by age, sex and population structure [29]. Recently, several studies have indicated that viral factors play a substantial role in SPVL variation by measuring its heritability between infections (reviewed in [30]). Most of these quantify the similarity in SPVL within transmission pairs, which are sexual couples in which one has infected the other [31–34].

The phenotype of any organism is controlled partly by its genome, and partly by its environment. Throughout this work we define heritability, h^2, in the broad sense as the proportion of total phenotypic variance (σ_P^2) ascribed to genetic variance (σ_G^2) [equation (1), [35]]. In the environmental component

of variance, σ_E^2, we conceptually include all host genetic and non-genetic effects, as well as interactions between host and virus genotype.

$$h^2 = \frac{\sigma_G^2}{\sigma_P^2}. \tag{1}$$

Alizon *et al.* [36] used a phylogenetic comparative approach to identify phylogenetic signal as a measure of heritability, without requiring behavioural data. Phylogenetic signal is the extent to which individuals with similar traits can be observed to cluster together on the phylogeny. This approach has the advantage that any sample of well-characterized patients could be analysed in this way. However, the authors did not account for cofactors such as age and co-infections, which influence the SPVL and may cluster together on the phylogeny. It is also uncertain exactly how the quantity measured by the two methods used (Pagel's λ and Blomberg's *K*) should be interpreted in the context of heritability. We herein propose a new approach which also uses only the phylogenetic relationships to determine genetic proximity of viruses, but additionally links the results to true heritability, and in some cases allows inclusion of the effect of cofactors on SPVL.

The aims of this study were to evaluate the phylogenetic approach for estimating heritability, to compare the efficacy of the various statistics available for quantifying heritability on simulated and real data, to use these methods to confirm the presence of heritability in previously analysed data, and to measure heritability in a dataset from the Netherlands which has not been previously analysed for this purpose.

METHODOLOGY

Data

Data from Rakai, Uganda
The study population was enrolled in the Rakai Community Cohort Study in the rural Rakai District of south-western Uganda. The study methods for this cohort have been outlined elsewhere [34, 37, 38]. We used those individuals who were sampled at a single time point in April 1995 ($n = 332$). Characteristics of the cohort are shown in Supplementary Table S1. Serum samples were used to measure SPVL, with most individuals providing a single measurement.

The phylogenetic analysis was conducted in RAxML [39] using the General Time Reversible

substitution model [40], the Γ model of rate heterogeneity with four distinct rate categories and a proportion of invariant sites [41] (GTR+Γ+I), which had the best score when the alignment was analysed with ModelTest [42]. The alignment was also analysed using the Recombination Analysis Tool (RAT [43]), which found no apparent recombination within the studied genes (data not shown).

A Simian Immunodeficiency Virus (SIV) outgroup was included (accession number: AB177846.1). The phylogeny was constructed from either the *gp41* (*env*) or *p24* (*gag*) sequences (Supplementary Figs S1 and S2), which were analysed separately because they have different substitution rates [44]. We also performed the phylogenetic and heritability analysis on the subtype A and D sequences separately, due to apparent imbalance in the joint trees.

Data from Switzerland
The Swiss data were taken from the Swiss HIV Cohort Study and its integrated genotypic drug resistance database which has been described elsewhere [45, 46], and the specific selection has been analysed in previous work [36, 47, 48]. The phylogeny, constructed from the reverse transcriptase and protease components of the *pol* gene (Supplementary Fig. S3), had been inferred in this previous work [36] using PhyML [49] from subtype B infected individuals for whom at least three viral load measurements were available after primary infection and before anti-retroviral therapy ($n = 661$). Characteristics of the cohort are shown in Supplementary Table S2.

Note that PhyML was used for the Swiss data, compared with RAxML in the other two datasets. However, we expect the methods to produce very similar results as both use a hill-climbing algorithm, likelihood scores are well matched when using the same nucleotide substitution model, and there is no evidence for systematic differences in the results [50].

We analysed four subdivisions of the data as in previous work [36]. The entire dataset ('All') fit the Liberal definition for viral load variability, meaning that at least three consecutive viral load measurements within the asymptomatic window (6–36 months after first positive viral RNA) remained within a one-log band of one another. The 'Strict' subdivision included only those individuals for whom the measurements in the asymptomatic window all sat within the one-log band. Excluding all but men who have sex with men (MSM) led to the further subdivisions, 'MSM' and 'MSM Strict'.

Data from Netherlands
The study cohort was provided by the ATHENA national observational cohort (seroconverters from 1996 or onwards) and the Amsterdam Cohort Studies (seroconverters before 1996) [51]. We included only individuals infected with subtype B between 1985 and 2008, for whom appropriate genetic data and SPVL were available [52] ($n = 416$).

We define SPVL as the mean of all $\log_{10}$ viral load measurements taken 6–24 months after the midpoint between the last negative and the first positive diagnosis. The phylogeny was reconstructed using a sequence of length 2064 containing the same elements of the *pol* gene as for the Swiss data.

We excluded codons strongly associated with drug resistance mutations (for details, see Supplementary Fig. S4). Four subtype C individuals from the same cohort were used as the outgroup. The phylogenetic analysis was performed using ModelTest [42], RAxML [39] and RAT [43] in the same way as the Rakai cohort, which also identified the same nucleotide substitution model as appropriate, and no recombination was detected (data not shown).

As with the Swiss data, individuals were categorized as 'All', 'Strict', 'MSM' and 'MSM Strict'. Additionally, two further categories, 'MSM NL' and 'MSM Strict NL', were created from the 'MSM' groups, which excluded individuals not originating in the Netherlands in order to further reduce confounding factors.

Trees for simulations were read and manipulated using the *ape* package [53] in R [54], which was also used to plot the trees.

Methods for calculating heritability statistics

A phylogeny, reconstructed from genetic data, is an approximation of the transmission network. Phylogenetic signal is a measure of how well trait values at the tree tips match their relative positions on the phylogeny, and several established methods are available to quantify this signal in terms of a single statistic: the Mantel test [55]; Blomberg's independent contrasts, which give us the Blomberg's K and PICv (variance of phylogenetic independent contrasts) statistics [56]; Pagel's λ transformation [57]; and the Abouheif–Moran (AM) tests [58], of which there are five variants ('oriAbouheif', 'sumDD', 'nNodes', 'patristic', 'Abouheif', with the latter used by default). We also developed two new methods which allow control of cofactors, the

phylogenetic pairs (PP) method and the hierarchical clustering (HC) method, which are described briefly here.

The PP method identifies pairs of individuals on the tree which are each other's closest neighbour, and these are assumed to be transmission pairs. Analysis of variance (ANOVA) identifies the degree to which the transmission partner explains an individual's SPVL. Crucially, the ANOVA approach allows for the inclusion of cofactors. These are age, sex and genital ulcer disease in the Rakai dataset (Supplementary Table S1); age, sex and risk group in the Swiss dataset (Supplementary Table S2); and age, sex, risk group, region of origin and the type of assay used to measure viral load in the Netherlands dataset (Supplementary Table S3). This method also ignores individuals who are not part of a phylogenetic pair.

The HC method is similar but considers larger clusters of individuals identified on the phylogeny by a threshold branch length, and examines the amount of variance in SPVL explained by the cluster. Because there is no intuitive ideal cluster size, proportion included or number of clusters to use, the method integrates over the range of clustering distances.

All established and new methods are described in detail in the Supplementary data.

Randomization test

The significance of a test statistic can be measured as in [56] by comparing the statistic derived from the data with its distribution under no heritability, which is the null hypothesis. Randomizing the tips of the tree by randomly reallocating the tips scrambles any heritability signal. This is performed 1000 times, and the analysis is repeated for each. The proportion of randomized datasets that give a statistic higher than the true value is the one-tailed *P*-value for presence of heritability. When the randomization test was performed for the *PP* and *HC* statistics the cofactors remained with their corresponding SPVL data.

Method of simulating SPVL data on a known phylogeny

This method uses a simple algorithm to simulate evolution of a continuous trait on a known phylogeny for a given heritability. This method is similar to that used in previous work [36] and is a variation on the Ornstein–Uhlenbeck process [59, 60], which allows

Brownian motion to occur while constraining the distribution of the population.

During the simulation, each node on the tree is assigned a trait value in $\log_{10}$ SPVL, beginning from the root, which is assigned the mean of the true SPVL data. Each daughter node is given a SPVL value depending on that of its parent, and on the SPVL distribution in the whole population. The higher the heritability, the more the value depends on the parent.

The SPVL at each subsequent daughter node, V_D [equation (2)], is derived from the parent node, V_P, the value h^2, and the random variable M which is normally distributed according to the mean and variance of the population $\log_{10}$ SPVL (distribution 3).

$$V_D = h^2 V_P + M\sqrt{1 - (h^2)^2}, \qquad (2)$$

$$M \sim N(\bar{\mu}, \sigma_P^2). \qquad (3)$$

The value h^2 is therefore the regression slope between the trait values at a parent node, V_P, and a daughter node, V_D (or an index and secondary case). This has been demonstrated to be equal to the broad sense heritability, h^2 as defined in equation (1) [61], and its further implications are discussed in the Supplementary data. The result is a set of data at the tips of the tree, simulated with known heritability h^2, which is used for analysis. We have also explored an alternative method of simulation which allows for multiple transmission between nodes on the tree, which is also described in the Supplementary data.

Multiple hypothesis testing to estimating true heritability and confidence intervals

For each value of h^2 between 0 and 1, in increments of 0.01, 100 simulations are performed and the relevant phylogenetic comparative statistics is calculated. A hypothesis test for the particular h^2 is then performed with these 100 values. They are compared with the statistic calculated from the true data, and the proportion which is lower than the true statistic becomes the probability *P* that the data are consistent with that value of h^2. The values of h^2 which produce *P*-values closest to 0.025, 0.975 and 0.500 become the lower and upper 95% confidence bounds, and the median estimate of h^2, respectively.

A visual distribution of h^2 is estimated using approximate Bayesian computation (ABC), of which details are given in the Supplementary data.

Estimate power to detect an effect

To examine the ability of each statistic to detect non-null heritability, we simulated 100 datasets for each value of a range (0–1) of h^2 values in increments of 0.05. To visualize the relationship between h^2 and the statistic of interest, we calculated the mean and standard deviation of the phylogenetic comparative statistics calculated from these 100 simulations. In order to estimate the power of each statistic to detect a significant effect at each value of h^2, we performed a randomization test on each simulated dataset, as described above (see 'Randomization test'), but with 100 randomizations. For each simulated dataset, the proportion of randomization tests which find significant ($P < 0.05$) heritability represents the power to detect an effect at that value of h^2.

All methods are available in R code on request from the corresponding author.

RESULTS

Testing significance in previously analysed data

The significance of heritability from the randomization test for the Rakai and Swiss data is shown in Tables 1 and 2, respectively.

A consistent signal was not found in the Rakai data (Table 1), but in the Swiss data it was more so (Table 2), and was higher for the smaller subdivisions. We tested the possibility that high significance in the smallest group was an artefact of small tree size, but this was not the case (Supplementary Table S5). We also tested the robustness of the results to phylogenetic uncertainty and found that small levels of perturbation (3%) have little effect on the results, but as the level of perturbation increases, the signal weakens accordingly (Supplementary Table S6).

Estimates of heritability in previously analysed data

Tables 1 and 2 show the medians and confidence intervals of h^2 as estimated by multiple hypothesis testing (MHT). In the Swiss data, the confidence intervals exclude zero in several cases. In all of these instances, the randomization test was also significant. However, a significant randomization test did not always correspond to confidence intervals which excluded zero, because these are one- and two-tailed tests, respectively.

The distribution of true heritability was calculated by ABC using both PP (red) and Pagel's λ (blue) for each subdivision in the Swiss and Rakai datasets (Figs 1 and 2). In the Swiss figure, the MSM Strict subdivision exhibits the distribution of h^2 most removed from zero, which supports the results from the randomization test and lower confidence bounds (Table 2). The method using Pagel's λ exhibits higher heritability in the larger subdivisions.

Similarly in the Rakai cohort, the most significant result from the randomization test (Table 1) is the *p24* subdivision analysed under Pagel's λ, and this is reflected in the results from ABC, in which the distribution of estimated h^2 is highly positive. Interestingly, the PP method finds that the most significant distribution visually is instead from the *gp41* subdivision which is also confirmed by the randomization test.

We also tested an alternative method of simulation which accounted for differences in branch length and allowed multiple generations between two adjacent nodes. We applied this to the Rakai *p24* subtype A data (Supplementary Table S7), and the alternative gave higher estimates of h^2 but wider confidence intervals.

Heritability was also measured in the Rakai and Swiss data using a phylogenetic mixed model, which assumes that the trait is determined by independent viral and host effects [62]. These results are given in Supplementary Table S4.

Testing the sensitivity of each statistic

The sensitivity of each phylogenetic comparative statistic to simulated heritability was explored by visualizing the relationship between them, and measuring the power of the statistic to successfully detect an effect of that size, in the entire Swiss dataset and the MSM Strict subdivision (Figs 3 and 4, respectively). None of the statistics had substantial power to detect heritability below $h^2 = 0.4$ on any subdivision, but most had the power to detect an effect above that level. The Mantel test performed poorly throughout. Blomberg's *K* performed well in the MSM Strict subdivision, but was insensitive when applied to the entire Swiss dataset, with only values above $h^2 = 0.6$ being consistently detected. In the entire Swiss dataset, the AM statistic appeared the most sensitive, but in the MSM Strict subdivision its performance matched that of Pagel's λ.

We also tested the power of variants of the HC and AM tests compared with their default methods

Table 1. The statistics (*Z*), *P*-values from a randomization test, medians and confidence intervals of h^2 from MHT on the Rakai data

Type of test	gp41 (n=325)				Subtype A gp41 (n=46)				Subtype D gp41 (n=214)				p24 (n=327)				Subtype A p24 (n=47)				Subtype D p24 (n=217)			
	Z	P	h^2	95% CI	Z	P	h^2	95% CI	Z	P	h^2	95% CI	Z	P	h^2	95% CI	Z	P	h^2	95% CI	Z	P	h^2	95% CI
PP	0.490	0.050	0.75	(0.00, 0.84)	-0.095	0.619	0.17	(0.00, 0.74)	0.311	0.340	0.66	(0.00, 0.83)	-0.099	0.668	0.08	(0.00, 0.56)	-0.089	0.633	0.02	(0.00, 0.81)	0.358	0.125	0.64	(0.00, 0.83)
HC complete	0.024	0.149	n.d.	n.d.	-0.001	0.514	n.d.	n.d.	0.029	0.056	n.d.	n.d.	-0.006	0.695	n.d.	n.d.	0.002	0.464	n.d.	n.d.	0.010	0.188	n.d.	n.d.
HC single	0.003	0.350	n.d.	n.d.	0.003	0.261	n.d.	n.d.	0.001	0.411	n.d.	n.d.	-0.003	0.670	n.d.	n.d.	-0.005	0.736	n.d.	n.d.	0.000	0.407	n.d.	n.d.
HC average	0.016	0.181	n.d.	n.d.	0.002	0.428	n.d.	n.d.	0.022	0.056	n.d.	n.d.	-0.008	0.783	n.d.	n.d.	0.001	0.477	n.d.	n.d.	0.017	0.040	n.d.	n.d.
HC ward	0.037	0.111	0.49	(0.00, 0.67)	-0.011	0.653	0.07	(0.00, 0.68)	0.041	0.038	0.59	(0.00, 0.78)	-0.011	0.756	0.08	(0.00, 0.49)	-0.004	0.555	0.08	(0.00, 0.73)	0.011	0.244	0.44	(0.00, 0.66)
Mantel	-0.018	0.779	0.72	(0.00, 0.95)	0.174	0.034	0.56	(0.00, 0.80)	-0.007	0.547	0.02	(0.00, 0.95)	-0.006	0.569	0.57	(0.00, 0.99)	0.024	0.392	0.39	(0.00, 0.59)	0.016	0.362	0.92	(0.00, 0.96)
Blomberg *K*	0.001	0.705	0.09	(0.00, 0.59)	0.861	0.044	0.47	(0.00, 0.60)	0.001	0.737	0.06	(0.00, 0.54)	0.724	0.050	0.47	(0.00, 0.60)	0.452	0.160	0.43	(0.00, 0.58)	0.563	0.250	0.38	(0.00, 0.56)
PICv	3101	0.335	n.d.	n.d.	11.5	0.797	n.d.	n.d.	4101	0.340	n.d.	n.d.	15.6	0.858	n.d.	n.d.	27.6	0.578	n.d.	n.d.	16.2	0.758	n.d.	n.d.
Pagel λ	0.473	0.275	0.25	(0.00, 0.58)	0.624	0.051	0.50	(0.00, 0.61)	1.55	0.169	0.87	(0.00, 1.00)	0.700	0.011	0.51	(0.02, 0.61)	0.000	0.746	0.10	(0.00, 0.54)	0.000	0.424	0.31	(0.00, 0.55)
AM oriAbouheif	-0.012	0.640	n.d.	n.d.	-0.030	0.614	n.d.	n.d.	0.022	0.301	n.d.	n.d.	0.045	0.107	n.d.	n.d.	0.014	0.426	n.d.	n.d.	0.024	0.295	n.d.	n.d.
AM Abouheif	-0.013	0.613	0.04	(0.00, 0.41)	-0.113	0.842	0.15	(0.00, 0.48)	0.019	0.277	0.20	(0.00, 0.49)	0.035	0.152	0.26	(0.00, 0.51)	-0.041	0.562	0.08	(0.00, 0.56)	0.021	0.291	0.29	(0.00, 0.51)
AM sumDD	-0.003	0.525	n.d.	n.d.	-0.046	0.799	n.d.	n.d.	-0.001	0.212	n.d.	n.d.	0.001	0.186	n.d.	n.d.	-0.031	0.629	n.d.	n.d.	-0.002	0.328	n.d.	n.d.
AM nNodes	-0.003	0.525	n.d.	n.d.	-0.046	0.799	n.d.	n.d.	-0.001	0.212	n.d.	n.d.	0.001	0.186	n.d.	n.d.	-0.031	0.629	n.d.	n.d.	-0.002	0.328	n.d.	n.d.
AM patristic	0.015	0.027	n.d.	n.d.	-0.021	0.356	n.d.	n.d.	0.020	0.030	n.d.	n.d.	-0.003	0.498	n.d.	n.d.	-0.018	0.368	n.d.	n.d.	-0.005	0.486	n.d.	n.d.

This table included from analysis of the *gp41* and *p24* genes in all the available Rakai data, or subtypes A or D separately. *P*-values showing borderline significance (*P* < 0.1) are in blue, and formal significance (*P* < 0.05) is in red, and confidence intervals in which the lower limit is above zero are also in red. n.d., not done.

Table 2. The statistics (Z), *P*-values from a randomization test, medians and confidence intervals of h^2 from MHT on the Swiss data

Type of test	All (n=661)				Strict (n=230)				MSM (n=404)				MSM Strict (n=134)			
	Z	P	h^2	95% CI	Z	P	h^2	95% CI	Z	P	h^2	95% CI	Z	P	h^2	95% CI
PP	−0.089	0.817	0.08	(0.00, 0.48)	0.376	0.020	0.61	(0.01, 0.77)	0.009	0.492	0.18	(0.00, 0.53)	0.595	0.005	0.68	(0.12, 0.82)
HC complete	0.008	0.214	n.d.	n.d.	0.045	0.038	n.d.	n.d.	0.020	0.096	n.d.	n.d.	0.049	0.032	n.d.	n.d.
HC single	0.006	0.159	n.d.	n.d.	0.030	0.038	n.d.	n.d.	0.008	0.175	n.d.	n.d.	0.034	0.040	n.d.	n.d.
HC average	0.004	0.315	n.d.	n.d.	0.039	0.043	n.d.	n.d.	0.016	0.113	n.d.	n.d.	0.046	0.024	n.d.	n.d.
HC ward	0.008	0.315	0.36	(0.00, 0.52)	0.060	0.020	0.58	(0.28, 0.72)	0.036	0.036	0.44	(0.00, 0.61)	0.063	0.023	0.62	(0.34, 0.73)
Mantel	−0.014	0.712	0.00	(0.00, 0.80)	0.040	0.201	0.91	(0.00, 0.97)	−0.005	0.516	0.00	(0.00, 0.76)	0.050	0.207	0.79	(0.00, 0.91)
Blomberg *K*	0.002	0.305	0.50	(0.00, 0.67)	0.025	0.366	0.44	(0.00, 0.72)	0.091	0.038	0.71	(0.00, 0.82)	0.593	0.080	0.42	(0.00, 0.54)
PICv	1594	0.307	n.d.	n.d.	127	0.345	n.d.	n.d.	40.2	0.052	n.d.	n.d.	5.58	0.002	n.d.	n.d.
Pagel λ	0.105	0.016	0.44	(0.16, 0.57)	0.216	0.030	0.44	(0.00, 0.57)	0.153	0.025	0.45	(0.00, 0.59)	0.646	0.016	0.54	(0.32, 0.63)
AM oriAbouheif	−0.013	0.680	n.d.	n.d.	0.016	0.361	n.d.	n.d.	0.038	0.141	n.d.	n.d.	0.127	0.023	n.d.	n.d.
AM Abouheif	−0.014	0.671	0.01	(0.00, 0.32)	0.015	0.338	0.24	(0.00, 0.49)	0.037	0.138	0.31	(0.00, 0.48)	0.126	0.019	0.49	(0.03, 0.61)
AM sumDD	0.000	0.166	n.d.	n.d.	0.001	0.168	n.d.	n.d.	0.001	0.136	n.d.	n.d.	0.013	0.016	n.d.	n.d.
AM nNodes	0.000	0.166	n.d.	n.d.	0.001	0.168	n.d.	n.d.	0.001	0.136	n.d.	n.d.	0.013	0.016	n.d.	n.d.
AM patristic	0.002	0.320	n.d.	n.d.	0.002	0.155	n.d.	n.d.	0.005	0.119	n.d.	n.d.	0.000	0.054	n.d.	n.d.

Individuals were subdivided according to variability of viral load (Strict), risk category (MSM) or both. *P*-values showing borderline significance ($P < 0.1$) are in blue, and formal significance ($P < 0.05$) is in red, and confidence intervals in which the lower limit is above zero are also in red. n.d., not done.

('ward' and 'Abouheif', respectively) (Supplementary Figs S5 and S6). The default methods performed equally well or better than any of the variations.

Application to data from the Netherlands

Supplementary Figure S4 shows the shape of the phylogeny inferred from the Netherlands data. These data have not been previously analysed for heritability, and so to reduce the problem of multiple testing only two methods, Pagel's λ and PP, were used to detect and measure heritability in each subdivision of the data.

Significant heritability was found only in one subdivision (MSM from the Netherlands with Strict viral loads) and using one statistic (Pagel's λ), which gave a heritability estimate of $h^2 = 0.52$ (Table 3). No effect was found when the PP method was used. None of the confidence intervals on h^2 excluded 0. The estimated distributions of h^2 are shown in Fig. 5.

DISCUSSION

In this study, we developed new methods for detecting and measuring heritability of SPVL using phylogenetic relationships and compared them with established methods on real and simulated data.

Which cohorts exhibit heritability?

Heritability was detected consistently in the MSM Strict subdivision of the Swiss cohort (Table 2), supporting the previous study of these data which found significant and high heritability in that subdivision [36]. Note that in that study, the level of significance was detected by randomizing the PICv rather than the *K* statistic, hence the strong result for PICv in the MSM Strict subdivision. We also found that several statistics uncovered significant heritability in all of the subdivisions of the Swiss data. In many such cases, the confidence intervals of h^2 also excluded zero, and the estimates for h^2 were high (0.44–0.68).
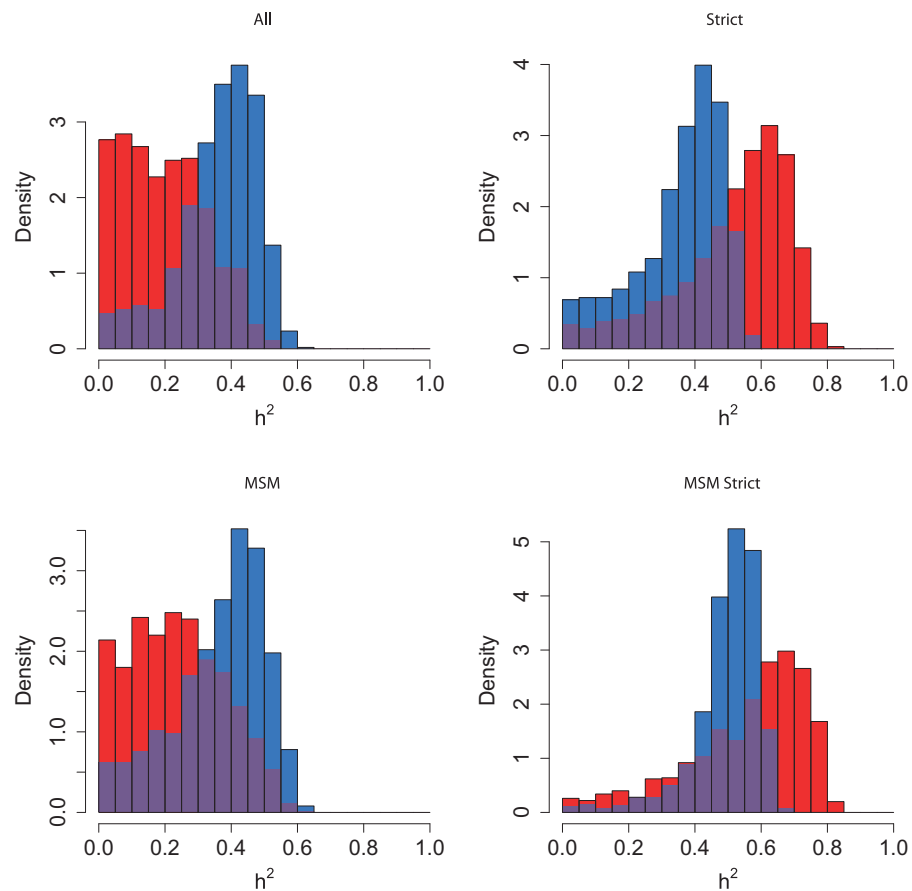
Figure 1. Heritability estimated by ABC in all subdivisions of the Swiss data. The subdivision is written above the plot. Results from the PP method are in red, and from Pagel's λ in blue, with the overlap in purple

In spite of significant heritability being found in previous work on the Rakai data [34], we did not consistently find heritability in this cohort. Although they are not strictly comparable, the results from the previous studies of the Rakai and Swiss data suggest that heritability is higher in the latter cohort, and this may underlie the differences in our results. However, the best estimate for h^2 from the Rakai data was high (0.51) (Table 1).

Six different subdivisions of the Netherlands data were analysed using PP and Pagel's λ, and only a single positive result was found with one statistic (λ) in one subdivision, suggesting that heritability is close to the detection borderline. It was estimated at $h^2 = 0.52$ (Table 3), which falls high within the range of previous estimates [30].

It is interesting that the Swiss cohort shows more apparent heritability than the Netherlands in spite of a similar transmission routes and genetic backgrounds. One potential explanation would be differences in coverage. The prevalence in Switzerland is higher (0.4% compared with 0.2%) [63], but so is the

sample size (661 compared with 416) which suggests similar levels of coverage, which are reportedly high in both cohorts [47, 64]. In the Swiss dataset, the mean SPVL of the Strict group is significantly higher than that of the rest ($P = 0.0004$), and the same is true of the MSM over non-MSM ($P = 0.003$) [36]. The same is not true of these two categories in the Netherlands cohort (data not shown), which suggests that viral virulence genotypes are less structured in this cohort and may explain lower heritability.

The exclusion of non-Strict individuals generally increased the level of significance, but this was dependent on the cohort and the method. In the Swiss cohort, the exclusion increased significance when using the PP and HC methods, but not otherwise (Table 2). In the Netherlands cohort, the exclusion increased significance markedly when using Pagel's λ, but not as much using PP (Table 3). Understanding why the Strict population exhibits higher heritability than whole sample may be an important step in both estimating and understanding the mechanisms behind heritability. Simple models
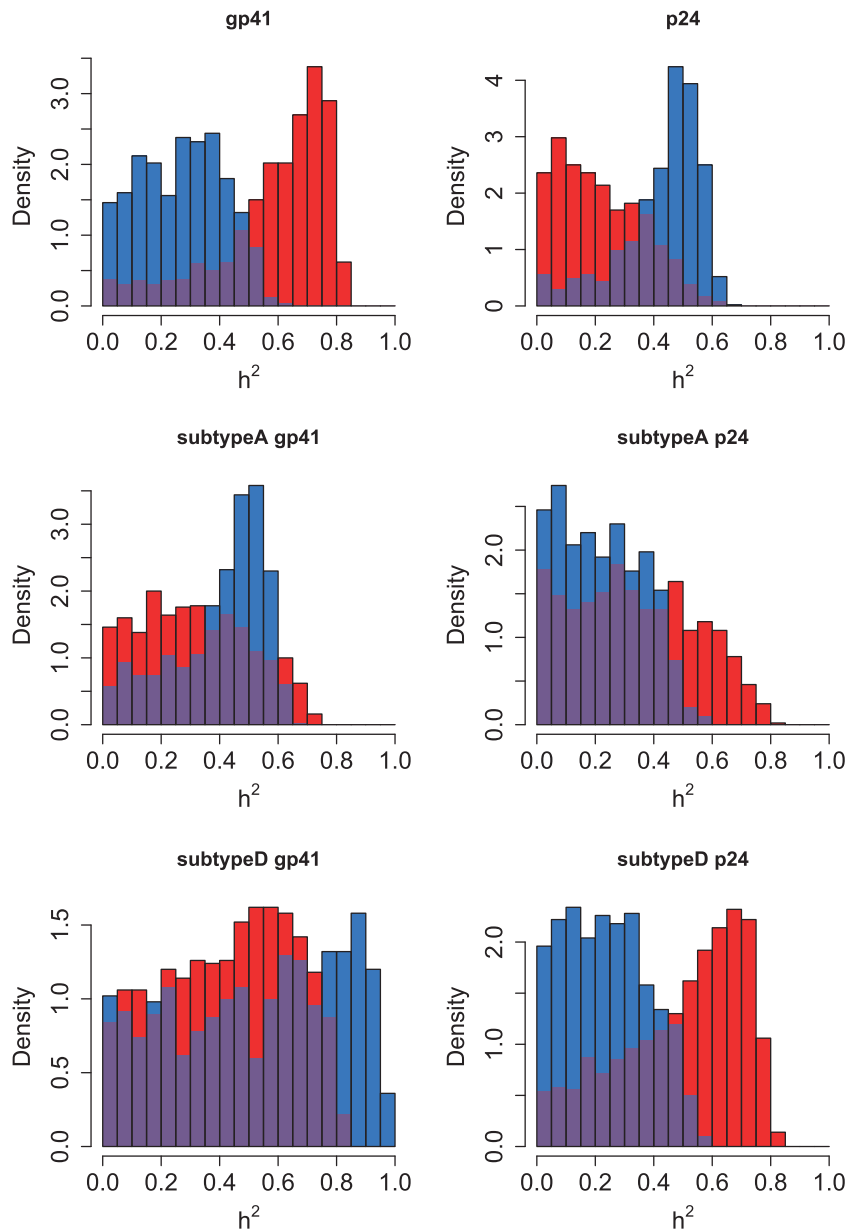
Figure 2. Heritability estimated by ABC in all subdivisions of the Rakai data. The subdivision is written above the plot. Results from the PP method are in red, and from Pagel's λ in blue, with the overlap in purple

fitted to longitudinal viral load data within patients suggest that fluctuations do not just represent random noise [65]. Fluctuating viral load has been associated with untreated sexually transmitted co-infections [66] and this or other host-mediated effect may mask the effect of the viral genotype and justify the exclusion of individuals with highly variable viral loads.

The same gene *pol* was used to analyse the Swiss and Netherlands datasets, whereas the available genes from the Rakai cohort were *env* and *gag*. It has been shown that *env* and *pol* produce similar trees [67], but there are topological differences, and our work also demonstrates different results between *gag* and *env* (Fig. 2). The ABC method of estimation by simulation makes estimation of h^2 robust to differences in gene usage and we do not expect systematic differences between the cohorts. There are also differences in transmission routes between the cohorts, and a higher diversity of HLA types in African than European populations [68]. Although these differences inhibit direct comparability between the European and Ugandan datasets, they add to the general nature of this work.
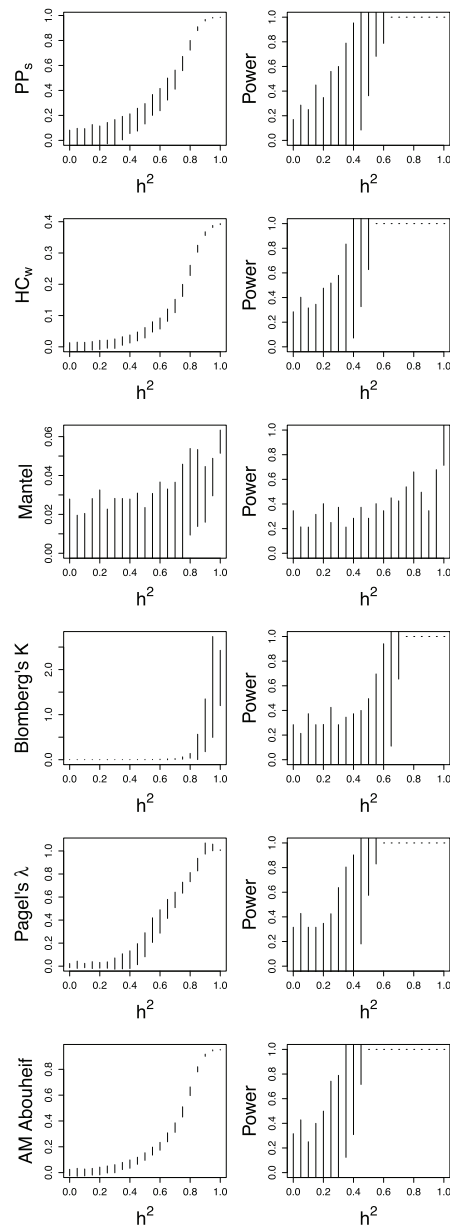
Figure 3. Comparison of the sensitivity of various statistics to heritability on the entire Swiss phylogeny. (Left side) The relationship between heritability and six different statistics under comparison. The circles and bars represent the mean and standard deviation of the sample. (Right side) The power of each statistic to detect heritability at 5% significance. The bars represent the standard deviation of the proportion

Figure 4. Comparison of the sensitivity of various statistics on the Swiss MSM Strict phylogeny. Left and right side as in Figure 3

Which method is best for detecting heritability?

The principle test of these statistics is the detection of an effect in real data. Pagel's λ detected heritability in every subdivision of the Swiss data (Table 2), and also produced the strongest result of any statistic applied to the Rakai dataset (Table 1). The PP and HC statistics performed well on the Swiss data, particularly on the Strict subdivisions. The AM statistics

were less successful at detecting an effect in the Swiss data. In the Rakai data, the AM 'patristic' variant found a significant result in two subdivisions. However, in simulation studies it performed very poorly (Supplementary Fig. S6).

Testing the detection power by simulation relies on a simple model of trait evolution, but has the advantage that heritability is known. It revealed that the PP, HC, Pagel's λ and AM statistics were comparably sensitive, detecting an effect at greater than approximately 0.4 heritability in the Swiss data, with

Table 3. The statistics (Z), P-values from a randomization test, medians and confidence intervals of h^2 from MHT on the Netherlands data, when analyzed with PP and Pagel's λ

	All	Strict	MSM	MSM Strict	MSM NL	MSM NL Strict
n	416	246	348	211	286	174
PP						
Z	−0.0627	0.2193	0.0763	0.2047	0.0633	0.2277
P-value	0.703	0.111	0.272	0.128	0.327	0.130
h^2	0.02	0.01	0.08	0.09	0.29	0.40
95% CI	(0.00, 0.45)	(0.00, 0.56)	(0.00, 0.53)	(0.00, 0.47)	(0.00, 0.63)	(0.00, 0.67)
Pagel's λ						
Z	0.0001	0.0978	0.0001	0.0323	0.0001	0.2628
P-value	0.824	0.086	0.827	0.200	0.718	0.038
h^2	0.19	0.44	0.01	0.38	0.10	0.52
95% CI	(0.00, 0.50)	(0.00, 0.60)	(0.00, 0.51)	(0.00, 0.60)	(0.00, 0.55)	(0.00, 0.63)

Individuals were subdivided according to variability of viral load (Strict), risk category (MSM) or whether they were from the Netherlands (NL). *P*-values showing borderline significance ($P < 0.1$) are in blue, and formal significance ($P < 0.05$) is in red, and confidence intervals in which the lower limit is above zero are also in red.

AM slightly more sensitive (Figs 3 and 4). The K statistic was as sensitive as other statistics when applied to the MSM Strict subdivision, but in the entire Swiss dataset its sensitivity was low.

Simulations suggest that none of the methods can detect heritability lower than approximately $h^2 = 0.4$, and this threshold is higher in some phylogenies. This threshold is confirmed by the finding that estimates of h^2 are always above 0.4 when heritability was found to be significant. Most studies have estimated lower heritability than this [30], and previous modelling work has suggested that such low heritability is enough to produce a substantial rate of evolution [15]. This suggests that phylogenetic methods are not adequate to exclude the possibility of relevant heritability in HIV virulence in these datasets.

Interestingly, another study which took an analytical and computational approach to comparing between Blomberg's K, the AM and the Mantel test found that K had a higher power to detect an effect than the AM statistic [69]. They also argue that these tests should all give the same significance as they are based on the cross-product of a phylogenetic similarity and trait similarity matrix. In contrast, we found marked differences between their performances, with the AM, K and Mantel statistic having decreasing power to detect an effect. They found that the sensitivity of these methods was dependent on the shape of the phylogeny, so differences in the source of trees (simulated versus inferred) are a possible source of discrepancy in the respective studies.

This is beyond the scope of this study but deserves to be the subject of future work.

The PP, along with the HC methods, is able to account for cofactors of SPVL: ignoring these may lead to overestimates or underestimates of heritability if they associate or dissociate on the tree, respectively. Indeed, an increase in signal when cofactors were taken into account was seen in previous analysis of transmission pairs in the Rakai cohort [34]. However, in the absence of a method to include the effect of cofactors in simulations, this aspect of the PP or HC methods cannot be harnessed to measure h^2. One possible method would be to calculate the fixed effects of the cofactors using a linear regression, simulate only the residuals, and subsequently add the fixed effects, but this requires treating residuals as data, which is inappropriate in the likely event that there is correlation between the effects of cofactors and the SPVL [70]. The PP method produces a dataset of couples, which is analogous to other couples studies [31–34]. The lack of sensitivity of the PP method and the other phylogenetic methods suggests that the phylogeny cannot (yet) tell us everything that the epidemiology does about the transmission network.

The λ statistic has the advantage that it incorporates both topology and branch lengths, and analyses the entire sample. It is notable, therefore, that the PP method is relatively successful in spite of its analysing only a subset of individuals who form apparent transmission pairs (60%), and in particular ignores deep relationships within the phylogeny.
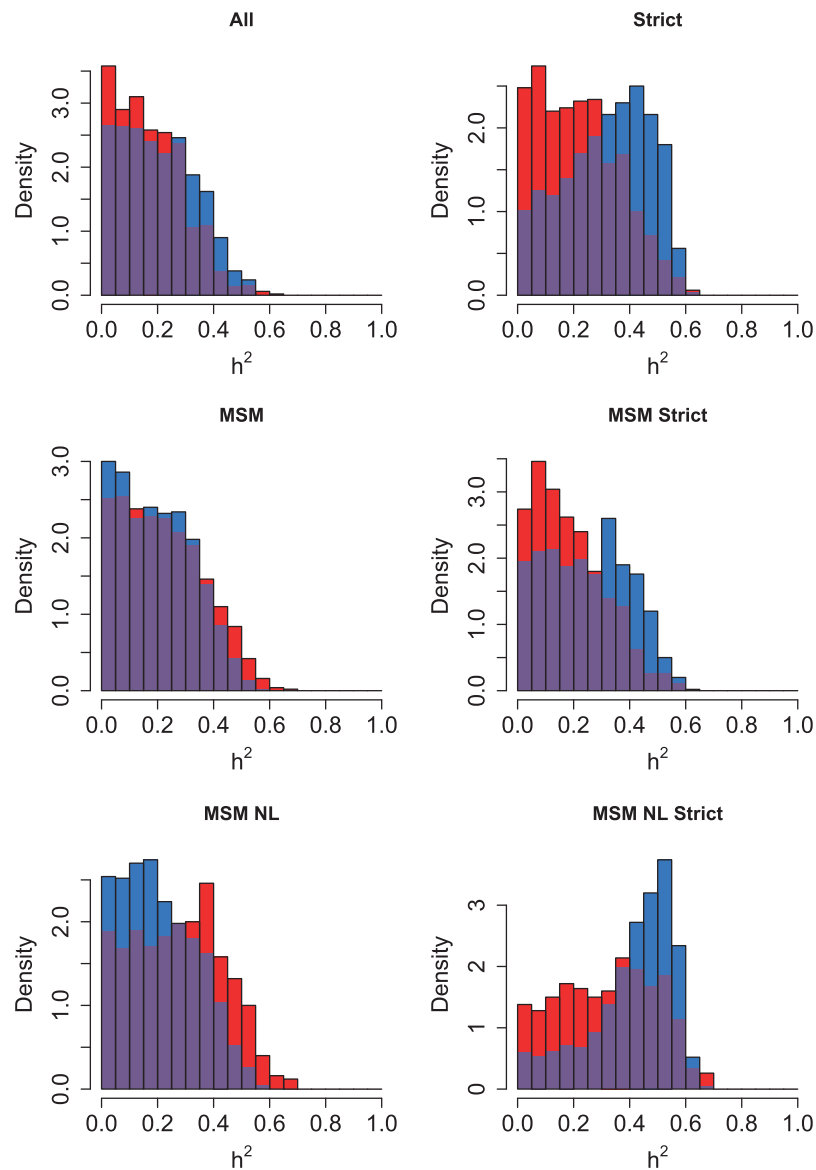
Figure 5. Heritability estimated by ABC in all subdivisions of the Netherlands data. The subdivision is written above the plot. Results from the PP method are in red, and from Pagel's λ in blue, with the overlap in purple

This suggests that most signal lies in the recent phylogenetic relationships. However, in unpublished work, Hodcroft *et al.* found SPVL heritability using pedigree analysis on UK data [71]. In contrast to our work, they found that collapsing poorly supported nodes in the tree and thereby ignoring some of the shallow relationships in the phylogeny had a negligible effect on their results for some datasets. The AM method is also successful, which (with the exception of the 'patristic' variant) ignores branch lengths, indicating that topology may be more important. Rigorously identifying which clades or levels of the phylogeny are responsible for heritability

would be an interesting direction for future research. This may differ for phylogenies which are less star-like than HIV. The use of longer sequences and better sampled datasets is likely to result in better detection and estimation of heritability, as poorly resolved trees scramble the heritability signal. However, the detection threshold is unlikely to change even with improved sampling, as it was based on simulations which were blind to uncertainty in the tree.

It is noteworthy that the PP and Pagel's λ each have their strengths in estimating the distributions of h^2 in the different Rakai subdivisions. The PP

method identifies a more strongly positive distribution in subdivision *gp41*, which Pagel's λ does not, but the latter detects a more positive distribution in *p24* and subtype A *gp41*. This suggests that an approach which combines these methods may be appropriate.

CONCLUSIONS AND IMPLICATIONS

In this study, we compare several phylogenetic comparative methods to detect heritability, h^2. Many methods detect heritability successfully in real and simulated data, but sensitivity drops off below $h^2 = 0.4$. We recommend the PP method and Pagel's λ for use in detecting and estimating heritability, the former for its consideration of co-factors, and the latter for its marginally higher level of sensitivity.

Estimates of heritability were consistent with previous studies on the Rakai and Swiss data, and confirm that heritability can be very high, which has clinical and evolutionary implications. When applied to the Netherlands data, heritability was found only in the most homogeneous subdivision, MSM who originate in the Netherlands with Strict viral loads. Differences in heritability between cohorts, subdivisions and methods for estimation carry implications for the biology of heritability, which offer interesting avenues for future modelling work. Experimental and epidemiological research are also required to directly identify viral factors which contribute to variance in SPVL, as well as exploring the impact of treatment on virulence evolution.

SUPPLEMENTARY DATA

Supplementary data are available at *EMPH* online.

ACKNOWLEDGEMENTS

Swiss HIV Cohort Study

Rakai Health Sciences Programme

ATHENA National Observational Cohort, The Netherlands

REFERENCES

1. Ji JP, Loeb LA. Fidelity of HIV-1 reverse-transcriptase copying a hypervariable region of the HIV-1 Env gene. *Virology* 1994;**199**:323–30.
2. Alizon S *et al*. Virulence evolution and the trade-off hypothesis: history, current state of affairs and the future. *J Evol Biol* 2009;**22**:245–59.
3. Ho DD *et al*. Rapid turnover of plasma virions and Cd4 lymphocytes in HIV-1 infection. *Nature* 1995;**373**:123–6.
4. Wei X *et al*. Viral dynamics in human immunodeficiency virus type 1 infection. *Nature* 1995;**373**:117–22.
5. Wolfs TFW *et al*. Naturally-occurring mutations within HIV-1 V3 genomic RNA lead to antigenic variation dependent on a single amino-acid substitution. *Virology* 1991;**185**:195–205.
6. Borrow P *et al*. Antiviral pressure exerted by HIV-1-specific cytotoxic T lymphocytes (CTLs) during primary infection demonstrated by rapid selection of CTL escape virus. *Nat Med* 1997;**3**:205–11.
7. Nowak MA *et al*. Antigenic diversity thresholds and the development of AIDS. *Science* 1991;**254**:963–9.
8. Richman D *et al*. Human-immunodeficiency-virus type-1 mutants resistant to nonnucleoside inhibitors of reverse-transcriptase arise in tissue-culture. *Proc Natl Acad Sci U S A* 1991;**88**:11241–5.
9. Schuurman R *et al*. Rapid changes in human-immunodeficiency-virus type-1 RNA load and appearance of drug-resistant virus populations in persons treated with lamivudine (3TC). *J Infect Dis* 1995;**171**:1411–9.
10. Leslie AJ *et al*. HIV evolution: CTL escape mutation and reversion after transmission. *Nat Med* 2004;**10**:282–9.
11. Poon AFY *et al*. Adaptation to human populations is revealed by within-host polymorphisms in HIV-1 and hepatitis C virus. *PLoS Pathog* 2007;**3**:e45.
12. Yerly S *et al*. Transmission of antiretroviral-drug-resistant HIV-1 variants. *Lancet* 1999;**354**:729–33.
13. Herbeck JT *et al*. Is the virulence of HIV changing? A meta-analysis of trends in prognostic markers of HIV disease progression and transmission. *AIDS* 2012;**26**:193–205.

14. Fraser C *et al*. Variation in HIV-1 set-point viral load: epidemiological analysis and an evolutionary hypothesis. *Proc Natl Acad Sci U S A* 2007;**104**:17441–6.
15. Shirreff G *et al*. Transmission selects for HIV-1 strains of intermediate virulence: a modelling approach. *PLoS Comput Biol* 2011;**7**:e1002185.
16. Dieckmann U *et al*. *Adaptive Dynamics of Infectious Diseases: In Pursuit of Virulence Management*. Cambridge: Cambridge University Press, 2002.
17. Gandon S *et al*. Imperfect vaccines and the evolution of pathogen virulence. *Nature* 2001;**414**:751–6.
18. Wolf F *et al*. AIDS prognosis based on HIV-1 RNA, CD4+ T-cell count and function: markers with reciprocal predictive value over time after seroconversion. *AIDS* 1997;**11**:1799–806.
19. Korenromp EL *et al*. Clinical prognostic value of RNA viral load and CD4 cell counts during untreated HIV-1 infection—a quantitative review. *PLoS One* 2009;**4**:e5950.
20. Mellors JW *et al*. Prognosis in HIV-1 infection predicted by the quantity of virus in plasma. *Science* 1996;**272**:1167–70.
21. Henrard DR *et al*. Natural history of HIV-1 cell-free viremia. *JAMA* 1995;**274**:554–8.
22. Geskus RB *et al*. The HIV RNA setpoint theory revisited. *Retrovirology* 2007;**4**:65.
23. Carrington M, O'Brien SJ. The influence of HLA genotype on AIDS. *Annu Rev Med* 2003;**54**:535–51.
24. Donnelly CA *et al*. Gender difference in HIV-1 RNA viral loads. *HIV Med* 2005;**6**:170–8.
25. Vüller M *et al*. Increasing clinical virulence in two decades of the Italian HIV epidemic. *Plos Pathog* 2009;**5**:e1000454.
26. Richardson BA *et al*. Comparison of human immunodeficiency virus type 1 viral loads in Kenyan women, men, and infants during primary and early infection. *J Virol* 2003;**77**:7120–3.
27. Abu-Raddad LJ, Patnaik P, Kublin JG. Dual infection with HIV and malaria fuels the spread of both diseases in sub-Saharan Africa. *Science* 2006;**314**:1603–6.
28. Jones LE, Perelson AS. Transient viremia, plasma viral load, and reservoir replenishment in HIV-Infected patients on antiretroviral therapy. *J AIDS* 2007;**45**:483–93.
29. Fellay J *et al*. Common genetic variation and the control of HIV-1 in humans. *Plos Genet.* 2009;**5**:e1000791.
30. Vüller M, Fraser C, Herbeck JT. A strong case for viral genetic factors in HIV virulence. *Viruses (Basel)* 2011;**3**:204–16.
31. Tang JM *et al*. HLA allele sharing and HIV type 1 viremia in seroconverting Zambians with known transmitting partners. *AIDS Res Hum Retroviruses* 2004;**20**:19–25.
32. Hecht FM *et al*. HIV RNA level in early infection is predicted by viral load in the transmission source. *AIDS* 2010;**24**:941–5.
33. van der Kuyl AC *et al*. HIV RNA levels in transmission sources only weakly predict plasma viral load in recipients. *AIDS* 2010;**24**:1607–8.
34. Hollingsworth TD *et al*. HIV-1 transmitting couples have similar viral load set-points in Rakai, Uganda. *Plos Pathog* 2010;**6**:e1000876.

35. Fisher RA. The correlation between relatives on the supposition of Mendelian inheritance. *Trans R Soc Edinb* 1918; **52**:399–433.

36. Alizon S *et al*. Phylogenetic approach reveals that virus genotype largely determines HIV set-point viral load. *Plos Pathog* 2010;**6**:e1001123.

37. Wawer MJ *et al*. A randomized, community trial of intensive sexually transmitted disease control for AIDS prevention, Rakai, Uganda. *AIDS* 1998;**12**:1211–25.

38. Wawer MJ *et al*. Control of sexually transmitted diseases for AIDS prevention in Uganda: a randomised community trial. *Lancet* 1999;**353**:525–35.

39. Stamatakis A. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* 2006;**22**:2688–90.

40. Tavaré S. Some Probabilistic and Statistical Problems in the Analysis of DNA Sequences. *American Mathematical Society: Lectures on Mathematics in the Life Sciences*, Vol. 17. American Mathematical Society, 1986, 57–86.

41. Yang ZH. Among-site rate variation and its impact on phylogenetic analyses. *Trends Ecol Evol* 1996;**11**:367–72.

42. Posada D, Crandall KA. MODELTEST: testing the model of DNA substitution. *Bioinformatics* 1998;**14**:817–8.

43. Etherington GJ, Dicks J, Roberts IN. Recombination Analysis Tool (RAT): a program for the high-throughput detection of recombination. *Bioinformatics* 2005;**21**:278–81.

44. Novitsky V *et al*. Intra-host evolutionary rates in HIV-1C env and gag during primary infection. *Infect Genet Evol* 2013;**19**:361–8.

45. Ledergerber B *et al*. The Swiss HIV Cohort Study—rationale, organization and selected base-line characteristics. *Soz Praventivmed* 1994;**39**:387–94.

46. Schoeni-Affolter F *et al*. Cohort profile: the Swiss HIV cohort study. *Int J Epidemiol* 2010;**39**:1179–89.

47. Kouyos RD *et al*. Molecular epidemiology reveals long-term changes in HIV type 1 subtype B transmission in Switzerland. *J Infect Dis* 2010;**201**:1488–97.

48. von Wyl V *et al*. Emergence of HIV-1 drug resistance in previously untreated patients initiating combination antiretroviral treatment: a comparison of different regimen types. *Arch Intern Med* 2007;**167**:1782–90.

49. Guindon S, Gascuel O. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol* 2003;**52**:696–704.

50. Sundberg K *et al*. *Incomparability of results between phylogenetic search programs*. http://www.ncbi.nlm.nih.gov/CBBresearch/Fellows/Carroll/pubs/sundberg2008incomparability.pdf (8 October 2013, date last accessed), 81–4.

51. Gras L *et al*. Viral load levels measured at set-point have risen over the last decade of the HIV epidemic in the Netherlands. *PLoS One* 2009;**4**:e7365.

52. Bezemer D *et al*. Transmission networks of HIV-1 among men having sex with men in the Netherlands. *AIDS* 2010; **24**:271–82.

53. Paradis E, Claude J, Strimmer K. APE: analyses of phylogenetics and evolution in R language. *Bioinformatics* 2004; **20**:289–90.

54. R Core Team. *R: A language and environment for statistical computing*. Vienna/Austria, http://www.R-project.org/ (8 October 2013, date last accessed).

55. Mantel N. The detection of disease clustering and a generalized regression approach. *Cancer Res* 1967;**27**: 209–20.

56. Blomberg SP, Garland T, Ives AR. Testing for phylogenetic signal in comparative data: behavioral traits are more labile. *Evolution* 2003;**57**:717–45.

57. Freckleton RP, Harvey PH, Pagel M. Phylogenetic analysis and comparative data: a test and review of evidence. *Am Nat* 2002;**160**:712–26.

58. Pavoine S *et al*. Testing for phylogenetic signal in phenotypic traits: new matrices of phylogenetic proximities. *Theor Popul Biol* 2008;**73**:79–91.

59. Uhlenbeck GE, Ornstein LS. On the theory of the Brownian motion. *Phys Rev* 1930;**36**:823–41.

60. Felsenstein J. Phylogenies and quantitative characters. *Annu Rev Ecol Syst* 1988;**19**:445–71.

61. Falconer DS. *Introduction to Quantitative Genetics*. Edinburgh/London: Oliver & Boyd, 1960.

62. Housworth EA, Martins EP, Lynch M. The phylogenetic mixed model. *Am Nat* 2004;**163**:84–96.

63. UNAIDS. *HIV and AIDS Estimates*. http://www.unaids.org/en/regionscountries/countries/ (8 October 2013, date last accessed).

64. Bezemer D *et al*. 27 years of the HIV epidemic amongst men having sex with men in the Netherlands: an in depth mathematical model-based analysis. *Epidemics* 2010;**2**: 66–79.

65. Masel J *et al*. Fluctuations in HIV-1 viral load are correlated to CD4(+) T-lymphocyte count during the natural course of infection. *J AIDS* 2000;**23**:375–9.

66. Atkins MC *et al*. Fluctuations of HIV load in semen of HIV positive patients with newly acquired sexually transmitted diseases. *Br Med J* 1996;**313**:341–2.

67. Rieder P *et al*. HIV-1 transmission after cessation of early antiretroviral therapy among men having sex with men. *AIDS (London, England)* 2010;**24**:1177–83.

68. Sanchez-Mazas A. African diversity from the HLA point of view: influence of genetic drift, geography, linguistics, and natural selection. *Hum Immunol* 2001;**62**:937–48.

69. Pavoine S, Ricotta C. Testing for phylogenetic signal in biological traits: the ubiquity of cross-product statistics. *Evolution* 2013;**67**:828–40.

70. Freckleton RP. The seven deadly sins of comparative analysis. *J Evol Biol* 2009;**22**:1367–75.

71. Hodcroft H *et al*. Population-based study indicates viral genetic effect on HIV virulence is small but significant, Conference on Retroviruses and Opportunistic Infections, Seattle, 2012.