

# Variation in HIV-1 set-point viral load: Epidemiological analysis and an evolutionary hypothesis

Christophe Fraser<sup>1,2</sup>, T. Déirdre Hollingsworth<sup>1</sup>, Ruth Chapman<sup>1,3</sup>, Frank de Wolf<sup>1,4</sup>, and William P. Hanage<sup>2</sup>

<sup>1</sup>Department of Infectious Disease Epidemiology, Faculty of Medicine, Imperial College London, London W2 1PG, United Kingdom; <sup>2</sup>Infectious Diseases Epidemiology Unit, London School of Hygiene and Tropical Medicine, London WC1E 7HT, United Kingdom; and <sup>3</sup>HIV Monitoring Foundation, Academic Medical Centre of the University of Amsterdam, 1105 AZ, Amsterdam, The Netherlands

Communicated by Stanley Falkow, Stanford University, Stanford, CA, September 10, 2007 (received for review February 9, 2007)

The natural course of HIV-1 infection is characterized by a high degree of heterogeneity in viral load, not just within patients over time, but also between patients, especially during the asymptomatic stage of infection. Asymptomatic, or set-point, viral load has been shown to correlate with both decreased time to AIDS and increased infectiousness. The aim of this study is to characterize the epidemiological impact of heterogeneity in set-point viral load. By analyzing two cohorts of untreated patients, we quantify the relationships between both viral load and infectiousness and the duration of the asymptomatic infectious period. We find that, because both the duration of infection and infectiousness determine the opportunities for the virus to be transmitted, this suggests a trade-off between these contributions to the overall transmission potential. Some public health implications of variation in set-point viral load are discussed. We observe that set-point viral loads are clustered around those that maximize the transmission potential, and this leads us to hypothesize that HIV-1 could have evolved to optimize its transmissibility, a form of adaptation to the human host population. We discuss how this evolutionary hypothesis can be tested, review the evidence available to date, and highlight directions for future research.

cohort studies | life-history | mathematical model | trade-off | virulence

Viral load, a measure of the density of virus particles in peripheral blood, is an imperfect but important measure of the severity of HIV-1 infection. Although its relationship to virus density in other body compartments and to viral replicative capacity is unclear (1–3), it has a proven track record in the prognosis of patients (4, 5) and has more recently been shown to predict the probability of transmission between discordant couples (6, 7). Viral load is heterogeneous both within patients over time and between patients. During the long asymptomatic period of infection, viral loads fluctuate around a steady set-point value, which varies up to 1,000-fold between patients (4, 5).

Although much work has considered the significance of primary infection in HIV-1 transmission (e.g., refs. 8 and 9), it is not known which set-point viral loads have the greatest epidemiological impact, in terms of leading to the greatest number of infections over the lifetime of the host. Such information is important to determine the public health consequences of targeting prevention efforts at patients with certain subsets of viral loads: high, low, or intermediate. Although it is clear that patients with higher viral loads will be more infectious (6, 7), it is also well known that, untreated, these individuals have a poorer prognosis and, hence, will have fewer lifetime opportunities for transmission (4, 5). The epidemiological impact of different viral loads will hence be determined by the interplay between these two antagonistic processes.

The aim of this study is to use available data on infectiousness (6, 7, 9) and duration of infection (5, 10) to determine, in general terms, the epidemiological impact of cross-sectional variation in viral load. We estimate the product of infectiousness and duration of infection, which we term the transmission potential. This is the mean number of persons one index case can potentially

infect over their whole asymptomatic period, estimated as a function of set-point viral load.

We quantify the transmission potential as a function of set-point viral load and find that it is maximized for intermediate viral loads, which we observe are also the most common among untreated patients. Although individuals with high viral loads are the most infectious in the short term, the total contribution to infection of those with intermediate viral loads is found to be larger because of the longer duration of asymptomatic infection. The consequences for public health and an evolutionary hypothesis arising from this observation are discussed.

## Results

### Set-Point Viral Load and the Duration of Asymptomatic Infection.

HIV-1 viral load follows a characteristic U-shape during the course of untreated infection, highest at the start (primary stage) and end (late stage) of infection, whereas lower and relatively steady levels are maintained for a variable number of years during asymptomatic infection. Levels measured in peripheral blood during the asymptomatic period are very variable, ranging from 1,000 to 1 million viral copies per milliliter, and this quantity is positively correlated with viral levels in other body compartments (1–5). We aim to study its relation to the duration of infection using a flexible parametric model in a robust inference framework.

To provide good quantitative detail, we focus on the Amsterdam seroconverters cohort, where homosexual men were recruited prospectively to study the incidence and natural history of HIV-1 infection from 1982 onwards and followed for many years; the cohort has been described elsewhere (5). We censor all observations after 22 November 1993, the date when the first protease inhibitor was used in this cohort, to avoid biases caused by the availability of effective treatment and ignore all treatment effects before this date; our sample size is 123 men, followed for 504 person-years. We determine set-point viral loads as the geometric mean viral load in the interval between the end of primary infection (defined as 6 months after first seropositive sample) and the first AIDS-defining event (CDC type C) or censoring, whichever occurs first. The distribution of these set-points is plotted in Fig. 1, as well as the distribution of viral loads collected from a Zambian cohort (7), referred to in more detail below.

We describe the duration of the asymptomatic stage of

Author contributions: C.F., F.d.W., and W.P.H. designed research; C.F., T.D.H., and R.C. performed research; C.F. contributed new reagents/analytic tools; C.F., T.D.H., and R.C. analyzed data; and C.F., T.D.H., R.C., and W.P.H. wrote the paper.

The authors declare no conflict of interest.

Freely available online through the PNAS open access option.

<sup>†</sup>To whom correspondence should be addressed at: Department of Infectious Disease Epidemiology, Faculty of Medicine, Imperial College London, St Mary's Campus, Norfolk Place, London W2 1PG, United Kingdom. E-mail: c.fraser@imperial.ac.uk.

This article contains supporting information online at [www.pnas.org/cgi/content/full/0708559104/DC1](http://www.pnas.org/cgi/content/full/0708559104/DC1).

© 2007 by The National Academy of Sciences of the USA





possible covariance between duration of infection and infectiousness, which would imply that the transmission potential, defined as the average of the product of duration and infectiousness, might not be well approximated by the product of the averages estimated here. Such a situation could arise because of joint dependence on cofactors other than viral load. Dependence on factors that independently affect survival or infectiousness, such as the dependence of infectiousness on sexual risk behavior, would not be problematic. Resolution of these concerns could be addressed by direct estimation of the transmission potential within a single patient cohort.

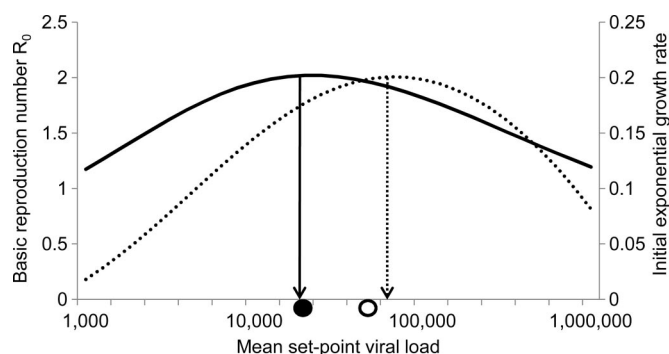
#### Consequences of Variation in Set-Point Viral Load for Public Health.

Comparison of Figs. 1 and 4 shows that individuals with common, intermediate viral loads have the largest transmission potential. For current public health initiatives based on the mass deployment of antiretroviral therapy, we suggest that to attempt to maximize indirect population benefits by singling out those with the highest viral loads for treatment would be misguided (a strategy explored but not advocated in ref. 13), because it is actually the majority of patients with intermediate viral loads who ultimately cause the most infections. Although indirect population benefits of mass therapy are possible and desirable, treatment protocols in areas of limited resource should use other inclusion criteria, such as clinical need, likelihood of treatment adherence, or sexual behavior.

As for future mass interventions based on, for example, imperfect vaccines, immunotherapy, or microbicides, this framework offers a simple tool for predicting so-called “perverse outcomes” (14, 15). If the intervention reduces patients’ viral load in such a way as to increase their transmission potential on average, then incidence will increase, not decrease. An intervention that reduces viral loads from high to intermediate levels and is therefore beneficial to the individual may nevertheless increase overall incidence and thus cause more overall harm than benefit. The ultimate outcome of any intervention that changes the distribution of viral loads can be predicted by calculating the change in the mean transmission potential. These conclusions are based on the epidemiological analysis of variation in set-point viral load and are independent of the evolutionary discussion that follows.

**A Hypothesis: The Evolution of HIV-1 Virulence.** The viral load that maximizes the transmission potential is 4.52 log<sub>10</sub> copies per milliliter (Fig. 4), close to the observed means of 4.36 and 4.74 for the Dutch and Zambian cohorts, respectively (Fig. 1). Viral loads during the asymptomatic period are clustered around values that maximize the transmission potential of the virus. Is it possible that this is not coincidence but, rather, an outcome of natural selection acting on HIV-1 to maximize opportunities for onwards transmission? This would suggest that HIV-1 conforms to the classical adaptive virulence model: Seen from the perspective of the virus, a negative correlation between infectiousness and duration of infection could be interpreted as a trade-off between two viral life-history traits, with natural selection leading to an optimal balance in this trade-off (16, 17).

This adaptive virulence model for HIV-1 results in a number of clear predictions that could be regarded as tests of the hypothesis. First, for the hypothesis to be true, the observed distribution of viral loads needs to be consistent with an evolutionary interpretation of the life-history tradeoff in the transmission potential. Our analysis supports this. Second, the hypothesis predicts that set-point viral loads in transmitter and recipient will be correlated. If a trait is heritable, the conclusion that natural selection can act on it follows automatically. Conversely, if this is not the case, it is impossible for natural selection to act on a trait, no matter what its relationship to fitness might be.



**Fig. 5.** Optimal viral loads in a simple transmission model. The basic reproduction number  $R_0$  (solid line) and the initial epidemic exponential growth rate (dashed line) are estimated as functions of the mean set-point viral load of a hypothetical viral “strain.” The viral load values that maximize these quantities are shown as arrows, whereas the observed mean viral loads for the cohorts are shown as circles (open for Zambia and filled for Amsterdam Seroconverters).

No studies have satisfactorily addressed the question of heritability in viral load to date. What evidence there is, direct and indirect in terms of other indicators of heritability in viral traits, is reviewed in *Discussion in SI Text*. We leave this as an open, testable prediction generated by our study.

Although there is no consensus on the dominant mechanisms of HIV pathogenesis, attention has shifted from within-host viral evolution (e.g., ref. 18) to pathological host immune activation (e.g., refs. 19 and 20). Viral load has been implicated as a measure of viral replication, which itself may regulate the rate of cell destruction, and mathematical models can capture the relation between set-point load and duration of the incubation period of AIDS (e.g., ref. 21). The adaptive-virulence model does not negate the role of host immune or environmental interactions in driving progression to disease but, rather, predicts that viral genetic factors modulate this progression, leading to marginally more or less severe distributions of outcomes in populations.

More specifically, the model predicts that the viral population will eventually become dominated by the “strain” with the largest basic reproduction number  $R_0$ , defined as the number of individuals infected by a typical index case in a totally susceptible population (16, 17). However, in an emerging epidemic, some strains could initially spread faster before being replaced by others with higher  $R_0$  [so-called  $r$ -selection (22)]. We attempt to disentangle these processes as follows.

Consider a transmission model with multiple hypothetical viral strains. Each strain is characterized by a distribution of set-point viral loads (representing the effects of host, environmental and chance variability), but some strains have an inherited tendency to produce slightly higher or lower viral loads, on average. We use an “age-of-infection” framework (23) to calculate the basic reproduction number,  $R_0$ , and the initial exponential rate of spread,  $r_0$ , of strains characterized by their mean set-point viral load. The details of the calculation, of assumed sexual mixing, and sensitivity to the parameters are presented in *Methods in SI Text*.  $R_0$  is maximized for a mean set-point viral load of 4.34 log<sub>10</sub> copies per milliliter, whereas the exponential growth rate  $r_0$  is maximized for a mean set-point viral load of 4.83 log<sub>10</sub> copies per milliliter. These predicted “optimal” values are close to the mean viral loads observed in the cohorts (Fig. 5).

Because both the negative relation between viral load and duration of infection and the positive relation between viral load and infectiousness can be understood in simple biological terms, the tradeoff between these in producing the peaked transmission potential curve does not, *per se*, suggest viral adaptation. How-



transmission. Although the best-fit curves looked different (with a higher Hill coefficient for female-to-male transmission), the model did not fit significantly better ( $P = 0.67$  based on likelihood ratio).

**Transmission Potential.** The transmission potential  $TP(V)$  is defined in the main text as the product of the infection rate during asymptomatic infection and the duration of asymptomatic in-

fection, i.e.,  $TP(V) = \beta(V)D(V)$ . Confidence intervals for the transmission potential were estimated as above.

We thank R. M. Anderson, N. M. Ferguson, N. C. Grassly, B. G. Spratt, and V. Müller for useful discussions. We gratefully acknowledge funding from the Wellcome Trust (R.C., T.D.H., F.d.W., and W.P.H.) and the Royal Society (C.F.) and the assistance provided by the Amsterdam Cohort Studies on HIV/AIDS.

- Hockett R, Kilby J, Derdeyn C, Saag M, Sillers M, Squires K, Chiz S, Nowak M, Shaw G, Bucy R (1999) *J Exp Med* 189:1545–1554.
- Coombs RW, Reichelderfer PS, Landay AL (2003) *AIDS* 17:455–480.
- Ball SC, Abraha A, Collins KR, Marozsan AJ, Baird H, Quinones-Mateu ME, Penn-Nicholson A, Murray M, Richard N, Lobritz M, *et al.* (2003) *J Virol* 77:1021–1038.
- Mellors J, Rinaldo C, Gupta P, White R, Todd J, Kingsley L (1996) *Science* 272:1167–1170.
- de Wolf F, Spijkerman I, Schellekens PT, Langendam M, Kuiken C, Bakker M, Roos M, Coutinho R, Miedema F, Goudsmit J (1997) *AIDS* 11:1799–1806.
- Quinn TC, Wawer M, Sewankambo N, Serwadda D, Li C, Wabwire-Mangen F, Meehan M, Lutalo T, Gray R (2000) *N Engl J Med* 342:921–929.
- Fideli OS, Allen SA, Musonda R, Trask S, Hahn BH, Weiss H, Mulenga J, Kasolo F, Vermund SH, Aldrovandi GM (2001) *AIDS Res Hum Retroviruses* 17:901–910.
- Jacquez JA, Koopman JS, Simon CP, Longini IM, Jr (1994) *J Acquir Immune Defic Syndr* 7:1169–1184.
- Wawer MJ, Gray RH, Sewankambo NK, Serwadda D, Li X, Laeyendecker O, Kiwanuka N, Kigozi G, Kiddugavu M, Lutalo T, *et al.* (2005) *J Infect Dis* 191:1403–1409.
- Lavreys L, Baeten JM, Chohan V, McClelland RS, Hassan WM, Richardson BA, Mandaliya K, Ndinya-Achola JO, Overbaugh J (2006) *Clin Infect Dis* 42:1333–1339.
- Morgan D, Mahe C, Mayanja B, Okongo JM, Lubega R, Whitworth JA (2002) *AIDS* 16:597–603.
- Morgan D, Mahe C, Mayanja B, Whitworth JAG (2002) *Br Med J* 324:193–196.
- Gray RH, Li XB, Wawer MJ, Gange SJ, Serwadda D, Sewankambo NK, Moore R, Wabwire-Mangen F, Lutalo T, Quinn TC (2003) *AIDS* 17:1941–1951.
- Anderson RM, Hanson M (2005) *J Infect Dis* 191:S85–96.
- Anderson RM, Gupta S, May RM (1991) *Nature* 350:356–359.
- Anderson RM, May RM (1982) *Parasitology* 85:411–426.
- Levin S, Pimentel D (1981) *Am Nat* 117:308–315.
- Nowak MA, Anderson RM, McLean AR, Wolfs TF, Goudsmit J, May RM (1991) *Science* 254:963–969.
- Brenchley JM, Price DA, Schacker TW, Asher TE, Silvestri G, Rao S, Kazzaz Z, Bornstein E, Lambotte O, Altmann D, *et al.* (2006) *Nat Med* 12:1365–1371.
- Silvestri G, Sooda DL, Koup RA, Paiardini M, O'Neil SP, McClure HM, Staprans SI, Feinberg MB (2003) *Immunity* 18:441–452.
- Fraser C, Ferguson NM, de Wolf F, Anderson RM (2001) *Proc R Soc Lond Ser B* 268:2085–2095.
- Pianka ER (1970) *Am Nat* 104:592–597.
- Levin BR, Bull JJ, Stewart FM (1996) *Math Biosci* 132:69–96.
- Levin BR, Bull JJ (1994) *Trends Microbiol* 2:76–81.
- Bonhoeffer S, Nowak MA (1994) *Proc Natl Acad Sci USA* 91:8062–8066.
- Gandon S, Mackinnon MJ, Nee S, Read AF (2001) *Nature* 414:751–756.
- O'Brien SJ, Nelson GW (2004) *Nat Genet* 36:565–574.
- Gray RH, Wawer MJ, Brookmeyer R, Sewankambo NK, Serwadda D, Wabwire-Mangen F, Lutalo T, Li XB, vanCott T, Quinn TC (2001) *Lancet* 357:1149–1153.