

Different Epidemic Potentials of the HIV-1B and C Subtypes

Marco Salemi,¹ Tulio de Oliveira,² Marcelo A. Soares,³ Oliver Pybus,⁴ Ana T. Dumans,⁵
Anne-Mieke Vandamme,⁶ Amilcar Tanuri,³ Sharon Cassol,² Walter M. Fitch¹

¹ Department of Ecology and Evolutionary Biology, University of California at Irvine, Irvine, CA 92664, USA

² Molecular Virology and Bioinformatics Unit, Africa Centre for Health and Population Studies, and the Nelson Mandela School of Medicine, Durban, South Africa

³ Laboratório de Virologia Molecular, Departamento de Genética, Instituto de Biologia, Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brazil

⁴ Department of Zoology, University of Oxford, Oxford, UK

⁵ Unidade de Genética, Universidade do Rio de Janeiro, Rio de Janeiro, Brazil

⁶ Rega Institute for Medical Research, KULeuven, Leuven, Belgium

Received: 4 July 2004 / Accepted: 11 October 2004 [Reviewing Editor: Dr. Yves Van de Peer]

Abstract. HIV, the cause of AIDS in humans, is characterized by great genetic heterogeneity. In particular, HIV-1 group M subtypes are responsible for most of the infections worldwide. We investigate the demographic history of HIV-1B and HIV-1C subtypes in South Africa and Brazil using both a parametric and a nonparametric approach based on coalescent theory. Our results show that although both subtypes are spreading exponentially in Brazil, the HIV-1C growth rate is about twice that of Brazilian HIV-1B or South African HIV-1C, providing evidence, for the first time, of a different epidemic potential between two HIV-1 subtypes. The present study not only may have important consequences for devising future vaccination and therapeutic strategies, but also offers additional evidence that skyline plots are indeed a simple and powerful tool for monitoring and predicting the behavior of viral epidemics.

Key words: Epidemic potential — HIV-1B — HIV-1C — Brazil — South Africa — AIDS

Introduction

HIV, the etiologic agent of AIDS, is classified into two distinct but clearly related types of viruses, HIV-1 and HIV-2, characterized by an extraordinary genetic variability (Hahn et al. 1984; Clavel et al. 1986). Three major groups can be distinguished within HIV-1: group M (for main), O (for outlier), and N (for neither, non-M–non-O, or new) (Robertson et al. 2000; Simon et al. 1998). Moreover, 9 phylogenetically distinct subtypes, 2 two subsubtypes, and at least 14 intersubtype HIV-1 recombinants, known as circulating recombinant forms (CRFs), have been repeatedly identified within group M so far (Robertson et al. 2000; McCutchan 2000; Salminen et al. 1995). Indeed, recombination coupled with the elevated error rate of the reverse transcriptase and the rapid turnover of HIV-1 in infected individuals are at the origin of the high genetic variability of the virus (Peeters and Sharp 2000).

Most subtypes, as well as CRFs, are present in Africa, reflecting the African origin of the epidemic (Gao et al. 1999). HIV-1B is the subtype responsible for most of the infections in Europe, United States, and Australia, whereas HIV-1C is the most prevalent worldwide, accounting for more than 56% of all infections (Esparza and Bhamarapravati 2000). It has been shown that the HIV-1 group M epidemic in

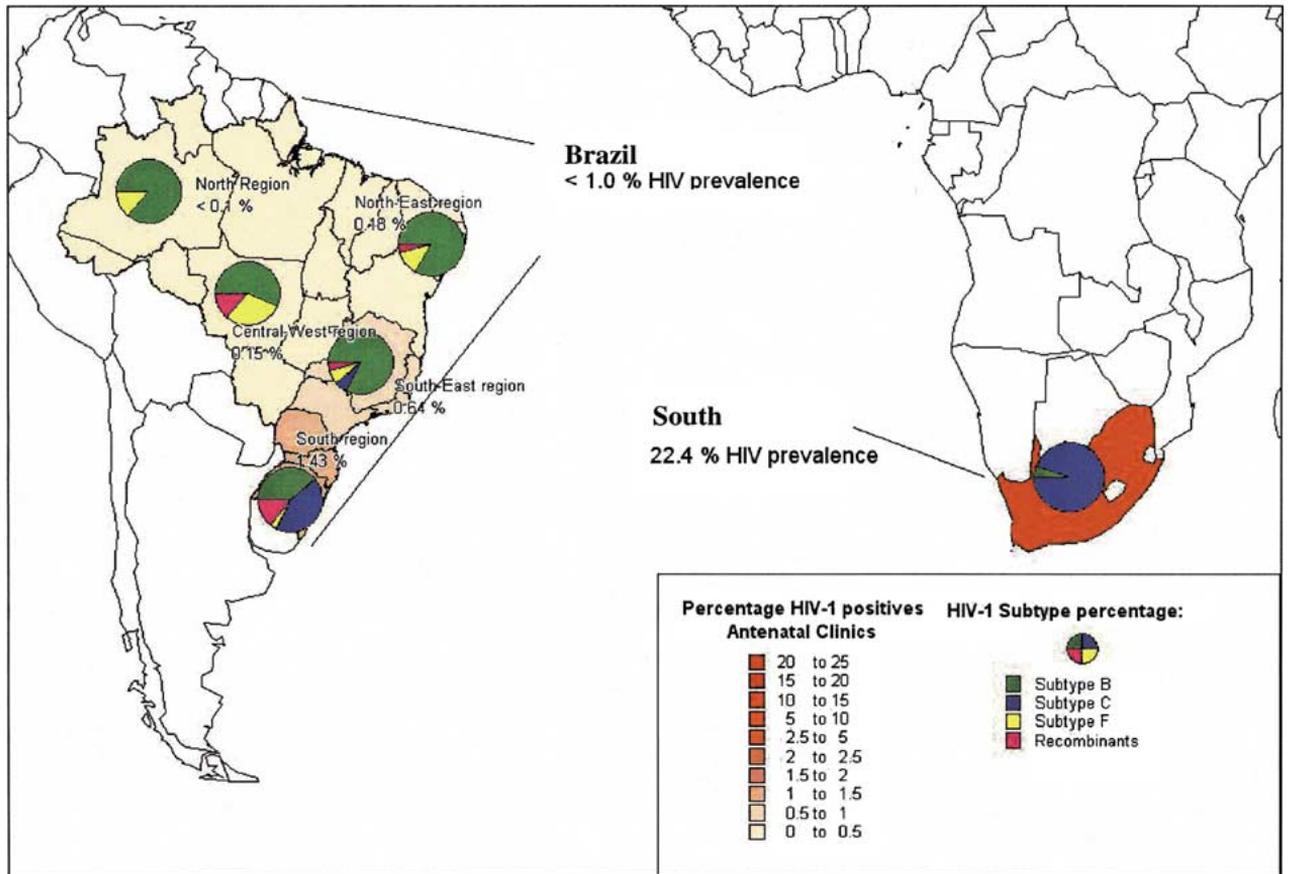


Fig. 1. Seroprevalence of different HIV-1 subtypes in South Africa and Brazil. The data for South Africa are based on a study of the South African National Health Department (HIV prevalence: HIV Antenatal Prevalence of the National Health Department 2001). The data for Brazil are based on a study by the Brazilian Institute of Health (HIV prevalence: Brazilian Antenatal HIV Prevalence in Sentinel Sites in 2000).

humans originated from a zoonotic transmission (Gao et al. 1999) and that the ancestor (most recent common ancestor) of group M probably dates back to the 1930s (Korber et al. 2000; Salemi et al. 2001). The existence of genetic subtypes could be the result of certain viral strains being involved in extensive transmission chains in a given geographic area, the so-called “founder effect.” In this view, in spite of their genetic differences, HIV-1 subtypes would be biologically equivalent. It is also hypothesized that characteristics such as a higher or lower transmissibility and/or fitness could explain the success or failure of different subtypes in different regions (Bjorndal et al. 1999). No definitive answer has been reached so far. In addition, the ability of a virus to spread in a population could be related to specific transmission routes like sexual contact and injecting drug use (Salemi et al. 1999; Pybus et al. 2001).

In the present study we investigate the epidemic behavior of HIV-1B in Brazil and HIV-1C in Brazil and South Africa. These two countries represent an interesting model to explore the potential differences between the two subtypes. The epidemic of HIV-1C is

relatively old in South Africa, where it is the most prevalent subtype (Fig. 1), whereas HIV-1B is dominant in Brazil. However, HIV-1C has been introduced recently in Brazil, and within the last few years it has become the most common subtype in recently diagnosed individuals in the southern states (Fig. 1) (Soares et al. 2003).

Materials and Methods

Data Sets

Six data sets have been compiled using B and C sequences that were nonrecombinant in the RT and protease genes: HIV-1B and HIV-1C Brazilian strains collected between 1992 and 2001 and HIV-1C South African strains collected between 1990 and 2002 (Cornelissen et al. 1997; Dumans et al. 2002; Gordon et al. 2003; Soares et al. 2003). To reduce nonrandom sampling, sequences were excluded if they came from the same patient or from individuals known to be related by a direct transmission. The data sets are summarized in Table 1. Sequences were aligned with the Clustal algorithm (Thompson et al. 1994) followed by manual editing. Positions containing gaps were removed from the final alignment. The alignments are available from the authors upon request.

Table 1. Likelihood ratio test comparing the likelihood of different demographic models for the HIV data sets

Isolate	Sequences	ln Lk1(exponential)	ln Lk2(logistic)	-2(Lk1 - Lk2)
Brazil				
HIV-1B				
Protease	83	-6.51917	-5.44127	2.1558 ($p = 0.142$)
RT	85	-33.0655	-32.9155	0.3 ($p = 0.584$)
HIV-1C				
Protease	51	-33.6105	-33.3943	0.4324 ($p = 0.511$)
RT	40	-13.4828	-12.3066	2.3524 ($p = 0.125$)
South Africa				
HIV-1C				
Protease	61	-26.0601	-24.9511	2.218 ($p = 0.136$)
RT	61	-2.97552	-1.86649	2.218 ($p = 0.136$)

Phylogeny and Evolutionary Rate Estimates

Maximum likelihood (ML) phylogenies were estimated for each data set. The best-fitting nucleotide substitution model was tested with a hierarchical likelihood ratio test following the strategy described by Swofford and Sullivan (2003), using a neighbor-joining tree with HKY85 estimated distances. ML phylogenies were then reestimated with the selected model, using a neighbor-joining tree as starting tree, and the TBR algorithm for branch swapping. Calculations were performed with PAUP* 4.0b10 (D.L. Swofford, Sinauer Associates, Sunderland, MA).

Because the sequences in our data set have been collected over several years, the evolutionary rate μ can be estimated via ML directly from the phylogenetic tree assuming a molecular clock with noncontemporaneous tips (Rambaut 2000). The molecular clock hypothesis can be tested with the likelihood ratio test with $n - 3$ degrees of freedom, where n is the number of taxa (Rambaut 2000).

Coalescent Models

By using coalescent theory, we can study how the ancestral relationships of individuals sampled from a population are influenced by its demographic history (Griffiths and Tavarè 1994; Kingman 1982).

A genealogy reconstructed from randomly sampled HIV sequences contains information about population-level processes such as change in population size and growth rate (Pybus et al. 2000). Given a viral phylogeny P and a vector ϕ representing the parameters of the model $N(t)$, it is possible to calculate the log of the conditional probability $\ln[\phi|P]$ (Pybus et al. 2000). ML estimates of ϕ can be found by numerical optimization of $\ln[\phi|P]$ and 95% CIs for the estimates obtained with the likelihood ratio statistic (Pybus et al. 2000). The estimated parameters are, in fact, $N(0)\mu$ and $r|\mu$, where μ is the evolutionary rate in nucleotide substitutions per site per year (the parameter c in the logistic model is unaffected by linear scaling of time). Notice that time runs backward into the past so that $N(0)$ is the effective number of the infections at the present, and $N(t)$ represents the effective number of infections at time t .

We consider two demographic models: exponential growth

$$N(t) = N(0)e^{rt} \quad (1)$$

and logistic growth

$$N(t) = N(0)(1+c)/(1+ce^{rt}) \quad (2)$$

where r is the growth rate achieved in a wholly susceptible population and c (Eq. [2]) is a logistic shape parameter related to the population carrying capacity (Pybus et al. 2001). Since the two models are nested they can be compared with the likelihood ratio test. The logistic model reduces to the exponential one when c is

infinite (degrees of freedom [df] = 1). In practice, c must be evaluated within a finite range and can be considered infinite when the upper limit of the 95% CI is 100,000 or more.

We also obtained nonparametric estimates of demographic history through the skyline plots (Pybus et al. 2000). However, the phylogenetic trees obtained for the HIV data sets show several zero or near-zero internal branch lengths, which make the skyline plots very noisy and more difficult to interpret. Therefore, we estimated the generalized skyline plots (Strimmer and Pybus 2001) for clock-like phylogenetic trees with dated tips. In such plots adjacent intervals smaller than a threshold of size ϵ in a tree are grouped together before obtaining the nonparametric estimates of the population size at any given time. For each data set, the optimal ϵ value used was the one maximizing the AICC (corrected Akaike information criterion) of the plot. All calculations were performed with GENIE version 3.0 (Pybus and Rambaut 2002).

Results

Evolutionary Rate Estimates

The estimated HIV-1 evolutionary rates for the different data sets were $1.40 \pm 0.28 \times 10^{-3}$ (HIV-B protease; Brazil), $1.54 \pm 0.38 \times 10^{-3}$ (HIV-B RT; Brazil), $3.38 \pm 0.95 \times 10^{-3}$ (HIV-C protease; Brazil), $3.55 \pm 0.64 \times 10^{-3}$ (HIV-C RT; Brazil), $1.71 \pm 0.27 \times 10^{-3}$ (HIV-C protease; South Africa), $1.47 \pm 0.21 \times 10^{-3}$ (HIV-C RT; South Africa). The molecular clock hypothesis was rejected by the likelihood ratio test for each data set. However, simulation studies have shown that if there is only a small amount of rate variation among lineages, then the 95% confidence limits of the rate estimate still contain the true mean rate about 95% of the time, even if the clock is rejected (Jenkins et al. 2002). In other words, the clock is very easily rejected, but the rate inferred enforcing the clock is still a good estimate of the mean rate and can still be used as a useful time scale.

Parametric Estimates of HIV-1B and 1C Demographic History in South Africa and Brazil

Using clock-like phylogenetic trees with noncontemporaneous tips and the estimated evolutionary rates, we tested the demographic hypotheses described by

Table 2. Estimates of the effective number of infections at present [$N(0)$], growth rate (r), and basic reproductive number (R_0) with an average duration of infectiousness D equal to 5 or 10 years for the different HIV data set (95% CIs in parentheses)

Isolates	$N(0)$	r	R_0 ($D = 5$ years)	R_0 ($D = 10$ years)
Brazil				
HIV-1B				
Protease	86,208 (40,463–190,146)	0.23 (0.20–0.26)	2.2 (2–2.3)	3.3 (3–3.6)
RT	107,111 (30,161–898,235)	0.35 (0.31–0.40)	2.8 (2.5–3)	4.5 (4.1–5)
HIV-1C				
Protease	2,543 (1,088–6,629)	0.77 (0.62–0.93)	4.9 (4.1–5.6)	8.7 (7.2–10.3)
RT	2,986 (623–8,599)	0.63 (0.51–0.75)	4.2 (3.6–4.8)	7.3 (6.1–8.5)
South Africa				
HIV-1C				
Protease	20,189 (7,567–60,032)	0.43 (0.36–0.51)	3.2 (2.8–3.6)	5.3 (3.7–6.1)
RT	19,071 (8,497–45,241)	0.28 (0.24–0.32)	2.4 (2.2–2.6)	3.8 (3.4–4.2)

Eqs. (1) and (2) for each of the data sets listed in Table 1. For none of them could the exponential model be rejected (see Table 1). Evidently, the number of infected individuals cannot grow forever and should eventually plateau according to the logistic model. At present, however, B and C subtypes in South Africa and Brazil appear to be spreading exponentially among the susceptible population, indicating that the population carrying capacity has not yet been reached. Table 2 shows the parametric estimates of $N(0)$ and r using the exponential model. For each HIV-1 data set similar results were obtained analyzing the protease or the RT gene. The confidence intervals, CIs, of $N(0)$ are quite large compared to those of r , but the estimates are in good agreement with the epidemiological data on the seroprevalence of subtypes B and C in Brazil and subtype C in South Africa (Department of Health, South Africa 2001; UN AIDS/WHO Working Group on Global HIV/AIDS and STD Surveillance 2000; see also Fig. 1). Note that $N(0)$ gives an estimate of the effective number of infections, i.e., those that are effectively transmitted in a population mating at random, not the absolute number of the infected people, which is usually larger.

The robustness of r estimates to change in $N(0)$ was tested by reestimating r while constraining $N(0)$ to vary over a range of values including its lower and upper 95% confidence limit given in Table 2 for each data set. In every case, the new estimates of r fell within the CIs of r reported in Table 2 (data not shown).

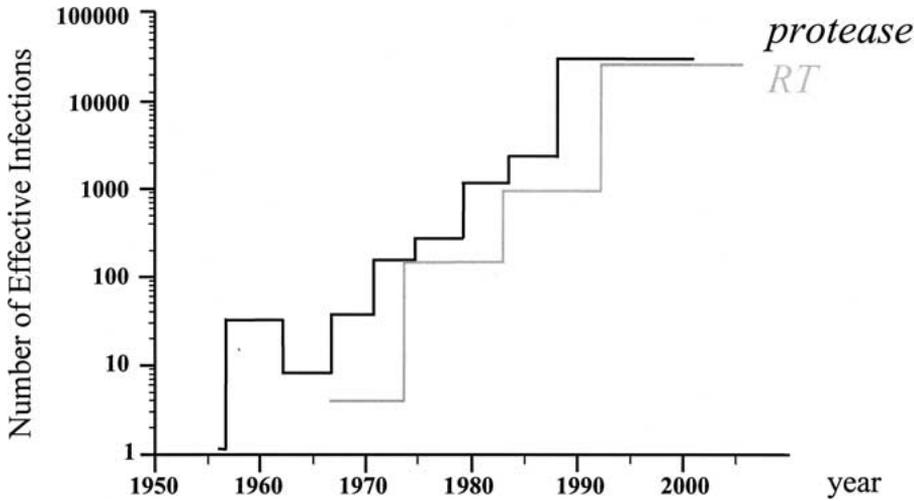
The r estimates also allow the estimation of the epidemiological quantity R_0 , the basic reproductive number (infectivity) of a pathogen, with the equation $R_0 = rD + 1$, where D is the average duration of infectiousness (Pybus et al. 2001). In Table 2 we use a putative but plausible range for D . For $D = 10$ years, on average eight secondary infections are generated by each primary HIV-1C infection in Brazil, versus about four secondary infections gen-

erated by HIV-1B in South Africa and Brazil or by the South African HIV-1C.

Generalized Skyline Plots of HIV-1B and 1C in South Africa and Brazil

Figure 2 shows the generalized skyline plots (Strimmer and Pybus 2001) with the nonparametric estimates of $N(t)$ for the different data sets. ML and nonparametric estimates of $N(0)$ appear to be in good agreement (see Table 2 and Fig. 2). According to the analysis of both protease and RT, the HIV-1B Brazilian epidemic appears to date back to between the middle of the 1950s and the middle of the 1960s, and probably originated, based on the phylogenetic analysis of subtype B and C Brazilian isolates (Soares et al. 2003), by one single introduction caused by one to three effective infections (see Fig. 2). The HIV-1C epidemics in South Africa and Brazil show a quite different story. The HIV-1-C South African epidemic originated sometime in the 1980s, with about three effective infections at its origin (see Fig. 2). The HIV-1C Brazilian epidemic is more recent (beginning of the 1990s), with its origin estimated to be a single effective infection (see Fig. 2). The data in Brazil are in accordance with a single recent introduction of the new subtype in the country (Soares et al. 2003). On the other hand, the skyline plot for HIV-1C in South Africa is consistent with phylogenetic analyses showing that the South African HIV-1C strains are not monophyletic and implying multiple separate introductions of the subtype around the beginning of the 1980s (Gordon et al. 2003). The plot also matches a report showing that the adult population infected in the country changed from 1% in 1990 to 10% in 1995 and to 24% in 2000 (Department of Health, South Africa 2001). By superimposing the protease skyline plots of HIV-1C in South Africa and Brazil (Fig. 3), it is clear that the Brazilian epidemic is spreading even faster and that the number of effective infections in Brazil might be expected to increase greatly within

Brazil HIV-1B



Brazil HIV-1C

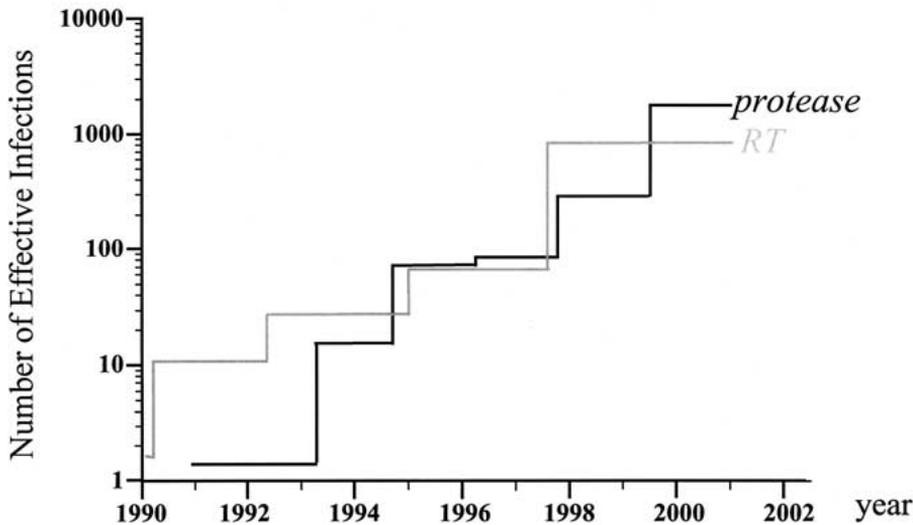


Fig. 2. Generalized skyline plots representing nonparametric estimates of $N(t)$ (Strimmer and Pybus 2001) using the protease (black line) and RT (gray or blue line) genes. The topmost step in the plot is the nonparametric estimate of $N(0)$. Genetic distances were transformed into a time scale of years using the evolutionary rates μ inferred for each data set. Since these plots are point estimates of $N(t)$, they do not contain information about uncertainty $N(0)$ and r (Table 2), or μ .

the next few years. RT skyline plots give similar results (data not shown).

Discussion

South Africa is one of the epicenters of the epidemic in the world, with about 22% of the adult population infected with HIV-1 (Department of Health, South Africa 2001), mostly with subtype C infections. Brazil, on other hand, is considered to be one of the best examples where therapeutic implementation has slowed down the course of the epidemic. Today less than 1% of the adult population is infected with HIV, about 600,000 people, mostly by subtype B and, more

recently, by subtype C in the southern region of the country (Dumans et al. 2002). HIV-1 growth rates are similar, except for HIV-1C in Brazil, which is spreading about two times faster than either HIV-1B or the South African HIV-1C. In particular, R_0 estimates in South Africa are in good agreement with current epidemiological data showing that each infected person has transmitted HIV to at least three new persons within 5 years (Department of Health, South Africa 2001). No evidence has been reported so far indicating that the average duration of infectiousness may vary among HIV-1 subtypes. Therefore, the twofold increase in infectivity of the Brazilian HIV-1C compared to the Brazilian HIV-1B and the South African HIV-1C may reflect a differ-

South Africa HIV-1C

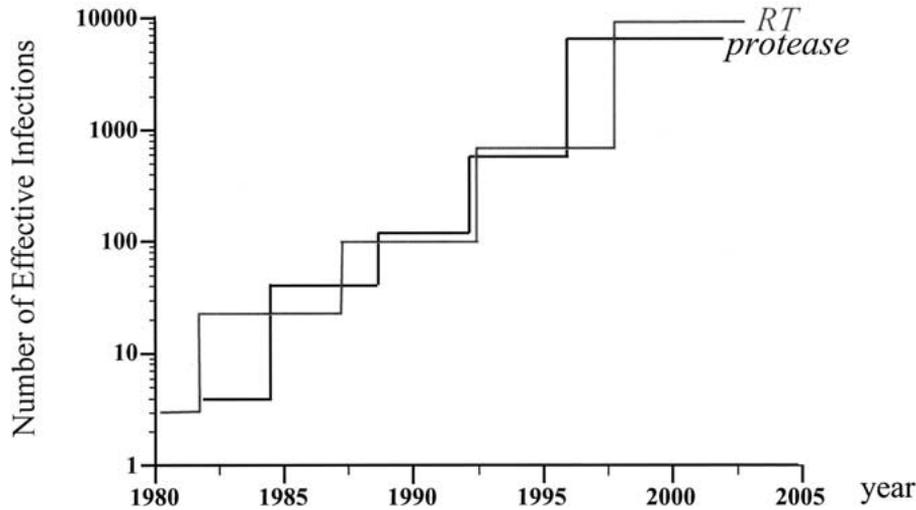


Fig. 2. Continued.

South Africa / Brazil HIV-1C *protease*

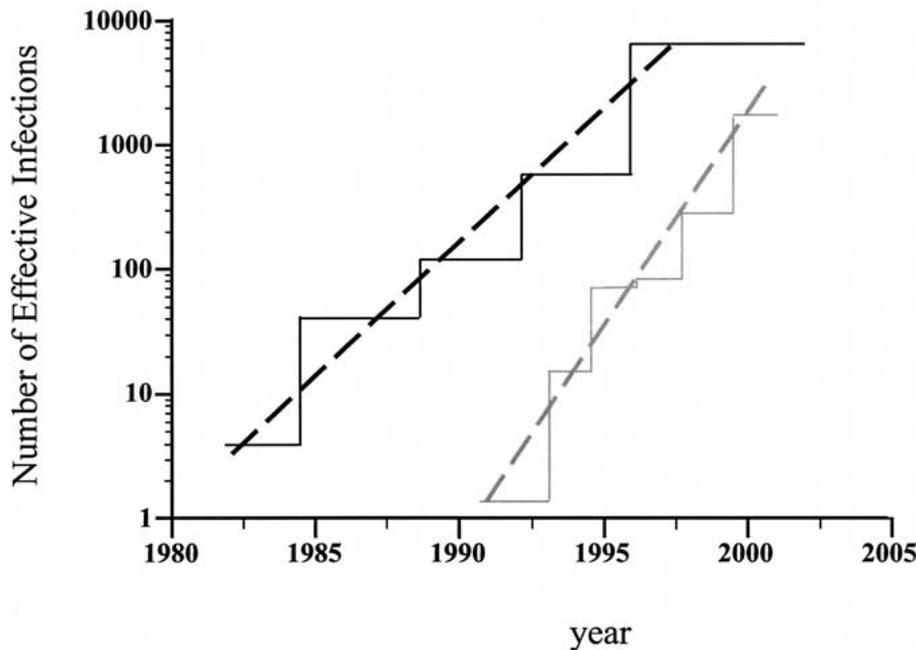


Fig. 3. Generalized skyline plots representing nonparametric estimates of $N(t)$ (Strimmer and Pybus 2001) using the protease gene for the South African (black line) and Brazilian (gray line) HIV-1C. The figure was obtained by superimposing the relative skyline plots reported in Fig. 1. Dashed lines represent the parametric estimates of $N(t)$ according to the exponential model (Table 2).

ence in the efficiency of different transmission routes in different geographic areas.

The above results depend on the assumptions of the coalescent model used: the evolutionary rate constancy, the absence of positive selection, recombination, and migration. We analyzed the protease and RT of naïve patients to reduce the effect of positive selection, and excluded recombinant strains,

but it is difficult to assess the importance of migration among subpopulations of infected HIV-1 patients in the countries studied. Also, the uncertainty in the evolutionary rate estimates may confound the interpretation of the analysis. However, the consistency of the results with current epidemiological data strengthens our confidence. Moreover, since our estimates of demographic history are consistent among genes, it

appears that the level of rate heterogeneity among HIV sequences is not large enough to systematically bias demographic inferences.

Overall we have shown that HIV-1C in Brazil is spreading at an increased rate. Following this trend the subtype may eventually become prevalent in the entire country, as has happened in the southern Brazilian states. Phylogenetic inference is too indirect to establish firmly whether the Brazilian HIV-1C is a new, more infectious strain or whether the virus is spreading faster because of a more favorable transmission route as suggested above. Yet the recently introduced HIV-1C is outcompeting HIV-1B in a country where the latter subtype was virtually the only one present until a few years ago, and a rapid escalation of HIV-1C infections has been occurring throughout sub-Saharan Africa, in India, and in China (Esparza and Bhamarapavati 2000; UNAIDS/WHO 2000). A similar scenario may be possible for other Western countries and deserves to be taken into account for future planning of vaccination and therapeutic campaigns around the world. These results also underline the need for refocusing prevention strategies in Brazil to stop the spreading of this viral variant. In this light, the use of viral gene sequences coupled with the results of coalescent theory appears to be a promising and important tool for monitoring and predicting the epidemic behavior of HIV subtypes and of other pathogens as well (Pybus et al. 2001; Robbins et al. 2003; Tanaka et al. 2002).

Acknowledgment. This work was supported by the Flemish Funds voor Wetenschappelijk Onderzoek (FWO Grants G.0288.01 and KAN2002 1.5.193.02, Postdoctoral Onderzoeker Contract 530).

References

- Bjorndal A, Sonnerborg A, Tscherning C, Albert J, Fenyo EM (1999) Phenotypic characteristics of human immunodeficiency virus type 1 subtype C isolates of Ethiopian AIDS patients. *AIDS Res Hum Retrovir* 15:647–653
- Clavel F, Guyader M, Guetard D, Salle M, Montagnier L, Alizon M (1986) Molecular cloning and polymorphism of the human immune deficiency virus type 2. *Nature* 324:691–695
- Cornelissen M, van den Burg R, Zorgdrager F, Lukashov V, Goudsmit J (1997) Pol gene diversity of five human immunodeficiency virus type 1 subtypes: evidence for naturally occurring mutations that contribute to drug resistance, limited recombination patterns, and common ancestry for subtypes B and D. *J Virol* 71:6348–6358
- Department of Health/Directorate Health Systems Research (2001) Seventh national HIV survey of woman attending antenatal clinics of the public health service of South Africa. October/November 2000. Directorate Health Systems Research. Department of Health, Pretoria, South Africa
- Dumas AT, Soares MA, Pieniazek D, Kalish ML, De Vroey V, Hertogs K, Tanuri A (2002) Prevalence of protease and reverse transcriptase drug resistance mutations over time in drug-naïve human immunodeficiency virus type 1-positive individuals in Rio de Janeiro, Brazil. *Antimicrob Agents Chemother* 46:3075–3079
- Esparza J, Bhamarapavati N (2000) Accelerating the development and future availability of HIV-1 vaccines: why, when, where, and how? *Lancet* 355:2061–2066
- Gao F, Bailes E, Robertson DL, Chen Y, Rodenburg CM, Michael SF, Cummins LB, Arthur LO, Peeters M, Shaw GM, Sharp PM, Hahn BH (1999) Origin of HIV-1 in the chimpanzee *Pan troglodytes troglodytes*. *Nature* 397:436–441
- Gordon M, De Oliveira T, Bishop K, Coovadia HM, Madurai L, Engelbrecht S, van Janse Rensburg E, Mosam A, Smith A, Cassol S (2003) Molecular characteristics of human immunodeficiency virus type 1 subtype C viruses from KwaZulu-Natal, South Africa: implications for vaccine and antiretroviral control strategies. *J Virol* 77:2587–2599
- Griffiths RC, Tavarè S (1994) Sampling theory for neutral alleles in a varying environment. *Philos Trans. R Soc London Ser B* 344:403–410
- Hahn BH, Shaw GM, Arya SK, Popovic M, Gallo RC, Wong-Staal F (1984) Molecular cloning and characterization of the HTLV-III virus associated with AIDS. *Nature* 312:166–169
- Jenkins GM, Rambaut A, Pybus OG, Holmes EC (2002) Rates of molecular evolution in RNA viruses: a quantitative phylogenetic analysis. *J Mol Evol* 54:156–165
- Kingman JFC (1982) The coalescent. *Stoch Proc Appl* 13:235–248
- Korber B, Muldoon M, Theiler J, Gao F, Gupta R, Lapedes A, Hahn BH, Wolinsky S, Bhattacharya T (2000) Timing the ancestor of the HIV-1 pandemic strains. *Science* 288:1789–1796
- McCutchan FE (2000) Understanding the genetic diversity of HIV-1. *AIDS* 14:S31–S44
- Peeters M, Sharp PM (2000) Genetic diversity of HTV-1: the moving target. *AIDS* 14 (Suppl 13):S129–S140
- Pybus OG, Rambaut A (2002) GENIE: estimating demographic history from molecular phylogenies. *Bioinformatics* 18:1404–1405
- Pybus OG, Rambaut A, Harvey PH (2000) An integrated framework for the inference of viral population history from reconstructed genealogies. *Genetics* 155:1429–1437
- Pybus OG, Charleston MA, Gupta S, Rambaut A, Holmes EC, Harvey PH (2001) The epidemic behavior of the hepatitis C virus. *Science* 292:2323–2325
- Rambaut A (2000) Estimating the rate of molecular evolution: incorporating non-contemporaneous sequences into maximum likelihood phylogenies. *Bioinformatics* 16:395–399
- Robbins KE, Lemey P, Pybus OG, Jaffe HW, Youngpairaj AS, Brown TM, Salemi M, Vandamme A-M, Kalish ML (2003) U.S. human immunodeficiency virus type 1 epidemic: date of origin, population history, and characterization of early strains. *J Virol* 77:6359–6366
- Robertson DL, Anderson JP, Bradac JA et al. (2000) HIV-1 nomenclature proposal. *Science* 288:55–56
- Salemi M, Lewis M, Egan JF, Hall WW, Desmyter J, Vandamme A-M (1999) Different population dynamics of human T cell lymphotropic virus type II in intravenous drug users compared with endemically infected tribes. *Proc Natl Acad Sci USA* 96:13253–13258
- Salemi M, Strimmer K, Hall WW, Duffy M, Delaporte E, Mboup S, Peeters M, Vandamme A-M (2001) Dating the common ancestor of SIVcpz and HIV-1 group M and the origin of HIV-1 subtypes using a new method to uncover clock-like molecular evolution. *FASEB J* 15:276–278
- Salminen MO, Carr JK, Burke DS, McCutchan FE (1995) Identification of breakpoints in intergenotypic recombinants of HIV type 1 by bootscanning. *AIDS Res Hum Retrovir* 11:1423–1425
- Simon F, Maucelere P, Roques P, Loussert-Ajaka I, Muller-Trutwin MC, Sargosti S, Georges-Courbot MC, Barre-Sinoussi F, Brun-

- Vezenet F (1998) Identification of a new human immunodeficiency virus type 1 distinct from group M and group O. *Nat Med* 4:1032–1037
- Soares MA, De Oliveira T, Brindeiro RM et al. (2003) A specific subtype C of human immunodeficiency virus type 1 circulates in Brazil. *AIDS* 17:11–21
- Strimmer K, Pybus OG (2001) Exploring the demographic history of DNA sequences using the generalized skyline plot. *Mol Biol Evol* 18:2298–2305
- Swofford DL, Sullivan J (2003) Phylogeny inference based on parsimony and other methods with PAUP*. In: Salemi M, Vandame A-M (eds) *The phylogenetic handbook—A practical approach to DNA and protein phylogeny*. Cambridge University Press, New York
- Tanaka Y, Hanada K, Mizokami M, Yeo AET, Wai-Kuo Shih J, Gojobori T, Alter HJ (2002) A comparison of the molecular clock of hepatitis C virus in the United States and Japan predicts that hepatocellular carcinoma incidence in the United States will increase over the next two decades. *Proc Natl Acad Sci USA* 99:15584–15589
- Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice. *Nucleic Acids Res* 22:4673–4680
- UN AIDS/WHO Working Group on Global HIV/AIDS and STD Surveillance (2000) Report on the global HIV/AIDS epidemic UN AIDS, Geneva