Ambiguous Nucleotide Calls From Population-based Sequencing of HIV-1 are a Marker for Viral Diversity and the Age of Infection

Roger D. Kouyos,1,a Viktor von Wyl,1,a Sabine Yerly,4 Jürg Böni,2 Philip Rieder,1 Beda Joos,1 Patrick Taffé,6 Cyril Shah,2 Philippe Bürgisser,7 Thomas Klimkait,8 Rainer Weber,1 Bernard Hirsche1, Matthias Cavassini,7 Andri Rauch,10 Manuel Battegay,9 Pietro L. Vernazza,11 Enos Bernasconi,12 Bruno Ledergerber,1 Sebastian Bonhoeffer,3 Huldrych F. Günthard,1 and the Swiss HIV Cohort Study

1Division of Infectious Diseases and Hospital Epidemiology, University Hospital Zurich, University of Zurich; and 2Swiss National Center for Retroviruses, University of Zurich, and 3Institute of Integrative Biology, Eidgenössische Technische Hochschule Zurich, Zurich; 4Laboratory of Virology, University Hospital Geneva and University of Geneva Medical School, and 5Division of Infectious Diseases, Geneva University Hospital, Geneva; 6Swiss HIV Cohort Study Data Center, and 7Division of Immunology, University Hospital Lausanne, Lausanne; 8Institute for Medical Microbiology, University of Basel, and 9Division of Infectious Diseases and Hospital Epidemiology, University Hospital Basel, Basel; 10Clinic for Infectious Diseases, Bern University Hospital and University of Bern; 11Division of Infectious Diseases, Cantonal Hospital St Gallen, St Gallen; and 12Division of Infectious Diseases, Regional Hospital Lugano, Lugano, Switzerland

Background. The time passed since the infection of a human immunodeficiency virus (HIV)–infected individual (the age of infection) is an important but often only poorly known quantity. We assessed whether the fraction of ambiguous nucleotides obtained from bulk sequencing as done for genotypic resistance testing can serve as a proxy of this parameter.

Methods. We correlated the age of infection and the fraction of ambiguous nucleotides in partial pol sequences of HIV-1 sampled before initiation of antiretroviral therapy (ART). Three groups of Swiss HIV Cohort Study participants were analyzed, for whom the age of infection was estimated on the basis of Bayesian back calculation (n = 3,307), seroconversion (n = 366), or diagnoses of primary HIV infection (n = 130). In addition, we studied 124 patients for whom longitudinal genotypic resistance testing was performed while they were still ART-naive.

Results. We found that the fraction of ambiguous nucleotides increased with the age of infection with a rate of .2% per year within the first 8 years but thereafter with a decreasing rate. We show that this pattern is consistent with population-genetic models for realistic parameters. Finally, we show that, in this highly representative population, a fraction of ambiguous nucleotides of .5% provides strong evidence against a recent infection event, 1 year prior to sampling (negative predictive value, 98.7%).

Conclusions. These findings show that the fraction of ambiguous nucleotides is a useful marker for the age of infection.

Human immunodeficiency virus type 1 (HIV-1) infections are initiated in most cases by a single virus [1], leading initially to a monomorphic viral population. Subsequently, viral diversity builds up gradually during HIV infection, first in a linear fashion but then at decreasing rates until a plateau is reached [2, 3]. In late-stage HIV infection, even decreases in viral diversity have been observed [3]. Thus, the diversity of the HIV population within an individual patient is potentially informative about the age of the infection, which is an important parameter because it allows an assessment of how far and how fast the infection has progressed. Seroconversion data are often lacking, and acute retroviral syndrome may have not occurred or may not have been recognized as such [4]. Thus, a method to estimate the
infection duration based on viral sequences would be attractive, given the abundance of HIV sequence data from genotypic drug resistance tests.

Genotypic resistance tests use nucleotide sequences to infer to what degree different drugs may inhibit a given viral population of an HIV-infected individual. For economic reasons, these sequences are obtained by bulk sequencing; that is, the sequencing procedure is applied to a diverse sample of the HIV population. If the frequency of the most frequent nucleotide at a given position exceeds a threshold (typically around 80%), bulk sequencing returns the predominant nucleotide at this position. However, if this is not the case, then so-called ambiguous nucleotide calls are reported, implying that the patient harbors viral strains with different nucleotides at this locus. Thus, the fraction of ambiguous nucleotides is a measure of the degree of polymorphism of the HIV population within a patient, which in turn should scale with the age of infection. Here, we assess to what degree the proportion of ambiguous nucleotides correlates with the time elapsed between HIV infection and sampling for genotyping. To this end, we relate the fraction of ambiguous nucleotides to the age of infection derived with methods of different accuracy from 4 data sets.

MATERIALS AND METHODS

We used previously analyzed nucleotide sequences from patients included in the Swiss HIV Cohort Study (SHCS) drug resistance database (see Kouyos et al [5] and GenBank accession numbers therein for a random sample of sequences). The SHCS is a nationwide, prospective, clinic-based cohort study with continuous enrollment and semiannual study visits [6, 7]. The SHCS has been approved by the ethical committees of all participating institutions, and written informed consent has been obtained from the participants. The SHCS drug resistance database contains the results of 13,201 genotypic resistance tests from 9,231 patients performed by the 4 laboratories engaged in HIV resistance testing in Switzerland, stored in a central database developed and hosted by SmartGene (Zug, Switzerland; Integrated Database Network System, version 3.5.0) [8]. Resistance data stem from routine clinical testing (60% of tests) and from tests performed retrospectively from frozen repository plasma samples (40% of tests). Retrospective sequencing was performed systematically by analyzing the earliest plasma sample available for each patient. All laboratories perform population-based sequencing of the full protease gene and at minimum codons 28–225 of the reverse transcriptase gene by means of commercial assays (Viroseq version 1, PE Biosystems; Virsoseq version 2, Abbott AG; vircotype HIV-1 assay, Virco Lab) and in-house methods [9].

Subtype B dominates the Swiss HIV-1 epidemic [10, 11], and we therefore focused on this subtype (n = 9,157 sequences). Because antiretroviral therapy strongly distorts viral diversity [12], we included only sequences from patients who were therapy-naive at the time of sampling and for whom an independent estimate of infection time was available (3,307 sequences, each from a different patient). The year of infection was estimated as described elsewhere [13] as the median of patient-specific infection time estimates based on a Bayesian back calculation model incorporating the dates of the first positive or last negative HIV test results and CD4 counts as predictor variables. Note that this infection time estimate is independent of the viral sequence. We refer to this data set as the full data set. Furthermore, we considered 3 additional sets of patients, for whom times or time differences are known with better accuracy: (1) In the set of seroconverters (366 patients), patients were recruited to the cohort within 1 year after infection, based on documented negative and positive HIV test results no longer than 180 d apart [11]; (2) The longitudinal set contains 124 patients for whom sequences are available at 2 time points (at least 6 months apart). Although for this data set the time of infection is not more precise than for the large data set, the age difference of the 2 samples is known exactly; and (3) The Zurich Primary HIV Infection Study (ZPHI; ClinicalTrials.gov identifier, NCT00537966) set contains 130 patients with sequences obtained during acute infection (median time from infection, 41 d; interquartile range [IQR], 28–56 d) [14].

The relationship between the proportion of ambiguous positions (the dependent variable) and the time since infection was investigated using linear regression analysis. Because model residuals were not normally distributed, we performed a bootstrap analysis with 1,000 replicates to obtain 95% confidence intervals (CIs). We verified results by repeating all analyses on logit-transformed fractions by use of a generalized linear model for proportions. Both crude and adjusted analyses, using the patient’s mode of HIV acquisition, ethnicity, sex, and age, were performed. In addition, a variable coding for the sequence-generating laboratory was included to account for assay- or laboratory-specific effects. All P values were 2-sided, and the level of statistical significance was set at .05.

We sought to identify a cutoff in the proportion of ambiguous sites to classify a patient’s infection status into recent (infected for ≤1 year) or chronic (infected for >1 year). The classification performance of 2 categorizations of the proportion of ambiguous sites was evaluated with receiver operating characteristic (ROC) analyses. The first analysis included ambiguous sites as a categorical variable with 5 groups based on quintiles, whereas for the second analysis only 2 categories were included based on an a priori defined cutoff of <=.5% or >.5% ambiguous positions.

RESULTS

In this study, we included HIV subtype B sequences of 3,307 patients. As shown in Table 1, the majority of individuals in our
Sample were male (n = 2,593; 78.4%) and of white ethnicity (n = 2,933; 88.7%). Of these 3,307 patients, 1,482 (45%) had acquired HIV through homosexual contacts, 862 (26.1%) through intravenous drug use, and 857 (25.9%) through heterosexual intercourse. The median year of infection was 1995 with an IQR of 1991–2000, and the median duration from infection to sampling for genotypic sequencing was 4.7 years (IQR, 3.3–6.9 years).

We found that for HIV-1 sequences sampled before initiation of antiretroviral therapy, the fraction of ambiguous nucleotides increases significantly with the age of infection. This relationship is shown in Figure 1A for the full data set (3,307 patients) for whom the age of infection has been estimated with the back calculation algorithm (see Materials and Methods and Taffe and May [13]). Figure 1A shows that the fraction of ambiguous sites grows linearly with time in the first few years of infection and

Table 1. Summary of the Adjusted Least Squares Regression for the Full Data Set

| Measure                                      | All patients | Patients infected ≤8 years |
|----------------------------------------------|--------------|----------------------------|
| Total no. of patients                        | 3,307 (100)  | 2,729 (100)                |
| Time since infection (per year)              | –            | –                          |
| Time since infection (quadratic term)        | –            | –                          |
| Female sex                                   | 714 (21.6)   | 542 (19.9)                 |
| Age by quartile, years                       |              |                            |
| 25–30                                        | 949 (28.7)   | 862 (31.6)                 |
| 33–41                                        | 789 (23.9)   | 605 (22.2)                 |
| 45–64                                        | 788 (23.8)   | 649 (23.8)                 |
| Ethnicity                                    |              |                            |
| White                                        | 2,933 (88.7) | 2,455 (90.0)               |
| Black                                        | 63 (1.9)     | 58 (2.1)                   |
| Hispanic-American                            | 97 (2.9)     | 89 (3.3)                   |
| Asian                                        | 57 (1.7)     | 55 (2.0)                   |
| Unknown ethnicity                            | 157 (4.7)    | 72 (2.6)                   |
| Mode of HIV acquisition                      |              |                            |
| Heterosexual risks                           | 857 (25.9)   | 722 (26.5)                 |
| Intravenous drug use                         | 862 (26.1)   | 595 (21.8)                 |
| Homosexual risks                             | 1,482 (44.8) | 1,326 (48.6)               |
| Unknown risk                                 | 106 (3.2)    | 86 (3.2)                   |
| Laboratory                                   |              |                            |
| Laboratory A                                 | 213 (6.4)    | 178 (6.5)                  |
| Laboratory B                                 | 1,450 (43.8) | 1,176 (43.1)               |
| Laboratory C                                 | 319 (9.6)    | 264 (9.7)                  |
| Laboratory D                                 | 1,325 (40.1) | 1,111 (40.7)               |
| Calendar year of sequencing, median IQR     | 2007 (2006–2008) | 2007 (2006–2008)        |
| HIV RNA load at time of GRT by 33 percentiles, log10 copies/mL |              |                            |
| 3.3–4.0                                      | 1,013 (30.6) | 829 (30.4)                 |
| 4.5–4.8                                      | 1,013 (30.6) | 859 (31.5)                 |
| 5.2–5.7                                      | 1,011 (30.6) | 853 (31.3)                 |
| HIV RNA load missing                         | 270 (8.2)    | 188 (6.9)                  |
| Constant term (1 unit = 1%)                  | –            | –                          |

**NOTE.** The 95% confidence intervals (CIs) were estimated with bootstrapping over 1,000 replicates. Note that for the quadratic model (all patients) both the linear and the quadratic term significantly improve the fit, but not when only the first 8 years of infection were considered. GRT, genotypic resistance test; HIV, human immunodeficiency virus; IQR, interquartile range; NA, not applicable.

* Statistically significant by the Wald test.
then flattens off. This effect can be described by including time since infection as a linear and a quadratic term in an adjusted regression model with the fraction of ambiguous nucleotides as outcome: if all patients are taken into account, both the positive linear and the negative quadratic components significantly improve the quality of the fit (model with linear term, $r^2 = .14$;
model with additional quadratic term, $r^2 = .17$) (Figure 1B; Table 1). Because the initial increase in the proportion of ambiguous positions is almost linear during the first phase of infection, we further restricted our analyses to patients with an HIV sequence obtained in this phase of linear increase, that is, within the first 8 years of infection ($n = 2,729$). These adjusted analyses revealed a yearly increase in the proportion of ambiguous sites of $\sim .2$ percentage points per year (Table 1), which was comparable with results from unadjusted regression analyses with yearly increases of $\sim .25\%$ (95% CI, 23%–28%) for the full sample and $\sim .19\%$ (95% CI, 18%–21%) for the sample restricted to patients with an infection duration of $\leq 8$ years. Interestingly, Table 1 indicates that the mode of HIV acquisition significantly affects the level of diversity. In particular, intravenous drug users exhibit a larger fraction of ambiguous nucleotides than do both heterosexuals and men who have sex with men, which may suggest broader transmission bottlenecks for blood-transmitted HIV.

To further assess the validity of the relationship between ambiguity and age of infection, we examined the smaller seroconverter set, which consists of 366 individuals who were recruited to the cohort in the early phase of their infection and for whom therefore the time point of infection was known within a given time interval of 6 months, and the longitudinal set, which contains 124 patients with sequence samples obtained from 2 time-points at least 6 months apart from each other. In the latter data set, inclusion was restricted to those patients for whom both samples were obtained within the first 8 years of infection. For these patients, we correlated the length of the time interval between the 2 measurements with the increase of the fraction of ambiguous nucleotides. Figure 1C shows that both additional data sets yield an estimate for the increase of ambiguous nucleotides over the first 8 years after infection that is similar to the estimate yielded by the original data set. Finally, we used the ZPHI data set, which contains the most stringent infection time estimate, to test the diversity around the time point of infection. Figure 1A shows that the fraction of ambiguous nucleotides in the ZPHI set is consistent with that found in the other 3 data sets. Moreover, for the ZPHI set, we observed a highly significant correlation (Spearman $p = 36; P < .001$) between the fraction of ambiguous nucleotides in the pol gene and the diversity of clonal env sequences [14], which confirms that the fraction of ambiguous nucleotides is a good marker for viral diversity.

In order to explore whether the observed correlation between infection time and the fraction of ambiguous sites could be exploited for age classification of individual HIV sequences, we subdivided our data set into quintiles of the fraction of ambiguous nucleotides $f$ and inferred the distribution of the time of infection for each of these quintiles. Figure 1D shows that the fraction of ambiguous nucleotides provides useful information mainly on the lower bound of the age of infection: whereas sequences with $f < .68\%$ (ie, the first two quintiles) (Table 2) may stem from the early phase of an infection, this is unlikely for sequences with a larger fraction of ambiguous nucleotides. However, it is important to note that although it is unlikely that the HIV sequence from a recent HIV infection has a large fraction of ambiguous nucleotides, this may occur if the HIV population is founded by $\geq 1$ virus, in which case diversity is large from the beginning on. This effect can be clearly seen in the ZPHI data set: a substantial minority of 24 (18%) of 130 patients exhibit a fraction of ambiguous nucleotides $> .68\%$, even though all patients' samples have been sequenced in the first months of their infection. This can be explained by results of Keele et al [1], who found that 23% of HIV infections are founded by $> 1$ virus and hence should exhibit a large diversity even during primary infection.

We explicitly evaluated 2 different stratifications for cutoffs to predict whether a patient was infected for $< 1$ year at the time of sampling for the genotypic test. As shown in Table 2, the selection of a cutoff of $\geq .5\%$ ambiguous positions predicted recent infections quite well in our full data set of 3,307 patients. Of 212 patients with a recent infection at the time of genotypic testing, 184 had $\leq .5\%$ ambiguous positions in their HIV sequence (sensitivity, 86.8%), but the specificity, which is the proportion of chronically infected patients with $> .5\%$ ambiguous positions, was only 70% (ROC area under curve, 78.3%). Therefore, this cutoff of $\leq .5\%$ may not be very useful to identify recently infected patients, but it can very accurately discriminate against patients with a chronic infection: only 28 (1.3%) of 2,190 patients with $> .5\%$ ambiguous positions had an infection duration of $\leq 1$ year at the time of genotyping. Thus, if the observed fraction of ambiguous positions is $> .5\%$, then there is a probability of 98.7% (95% CI 98.2%–99.1%) that the genotypic test was performed $> 1$ year after infection (negative predictive value). Stratification by quintile did not improve the classification performance but essentially confirmed the above finding: when quintiles 1 and 2 are collapsed into a single category and compared with the remaining strata (ie, sequences with $< .68\%$ vs those with $\geq .68\%$ ambiguous positions), sensitivity increased to 89.6% but specificity decreased to 63.2%. When testing the $.5\%$ cutoff in the sample of observed seroconverters ($n = 366$), the performance was very similar to that in the full sample. Sensitivity and specificity were 84.3% and 59.8%, respectively (data not shown). Overall, 79.8% of patients were correctly classified, and the area under the ROC curve was 72.0%.

The observed increase in ambiguous positions over the time of infection likely reflects the diversification of the HIV population after the genetic bottleneck at the infection event. In the simplest scenario, this diversification is controlled only by mutation and genetic drift (ie, random extinction of viral strains). We modeled this scenario by simulating the Wright-Fisher model (WFM) [15], starting from an initially uniform...
population (ie, with no initial diversity). As the WFM assumes neutrality, we focused on ambiguous nucleotides at 4-fold degenerate third-codon positions for which the assumption of neutrality is best justified. We found that the WFM can reproduce the temporal increase of the fraction of ambiguous sites for parameters that have been shown to reproduce neutral evolution in HIV (see Kouyos et al [16] and references therein) (Figure 2). This finding implies that, at least for 4-fold degenerate third-codon positions, the increasing diversity of the HIV population can be understood as the combined effect of mutation and genetic drift acting on a virus population, which is homogenous at the time point of infection.

**DISCUSSION**

The most relevant clinical result of this study is that a large fraction of ambiguous nucleotides provides evidence against a recent infection event. In particular, we found that a threshold of .5% ambiguous nucleotides yields a negative predictive value of 98.7%. It is important to note that, by definition, the negative predictive value depends on the overall composition of the considered population at time of diagnosis. A strength of the sample studied here, however, is the high representativeness of the SHCS for the epidemic in Switzerland [6–8]. Thus, the composition of our sample is likely to be very similar to that of the population of HIV-1-infected individuals at the time of HIV infection diagnosis, and therefore the negative predictive value inferred here closely describes the situation in clinical practice.

Finally, we find that although the frequency of ambiguous nucleotide calls differs significantly between laboratories (Table 1), the negative predictive value varies only marginally when calculated for samples from each laboratory separately (range, 98.5%–99.5%; $\chi^2$ test, $P = .668$).

In order to test the robustness of our findings, we additionally assessed the impact of HLA types, transmitted resistance, and subtype. HLA types may affect viral diversity, for instance, by stabilizing selection (and possibly by maintaining detrimental viral mutations, as may be the case for type HLA-B*57) or, in the opposite, by leading to increased viral replication and diversity by triggering only suboptimal cytotoxic T lymphocytes responses, for instance, in the presence of homozygous HLA alleles. Using available HLA data from 352 individuals, we considered the homozygosity of HLA-B ($n = 32$) and the following HLA-B haplotypes: HLA-B*57 ($n = 23$), HLA-B*27 ($n = 25$), HLA-B*5801 ($n = 7$), and HLA-B*35 ($n = 80$).
Upon inclusion of these variables in the adjusted regression model (restricted to a time of ≤8 years), neither of these HLA-B alleles reached statistical significance. There was a nonsignificant trend for higher viral diversity among carriers of homozygous HLA-B alleles of .243 (95% CI -.054 to .551). Concerning HIV subtype, it should be noted that the present analysis focused on subtype B, because most patients in the SHCS have been infected with that subtype. We found, however, very similar results for other subtypes (Table 3), indicating that our methodology extends beyond subtype B. However, it is also clear from the limited data available for non-B subtypes that additional tests are required for these subtypes. Finally, transmitted resistance mutations (present in 10% of the included patients) did not seem to have an impact when included in an adjusted model analogous to that in Table 1 (with time restricted to ≤8 years), since the parameters were not statistically significant (data not shown). In summary, these 3 tests suggest that our findings are robust against HLA type, subtype, and transmitted resistance.

A potential limitation of our analysis is the accuracy of infection date estimates. We have confirmed the correlation between the age of an infection and the fraction of ambiguous base calls in several data sets with different estimation methods for infection duration. Moreover, we performed a sensitivity analysis to assess the impact of the uncertainty of infection date estimates from the back calculation model (data not shown). This was done by repeating the adjusted and unadjusted regression analyses on a random time point between the upper and the lower bound of the posterior infection date distribution [13]. Time trends were somewhat attenuated due to the introduction of additional variation with an increase of .13% per year (95% CI, .11%–.14%) for the unadjusted analysis and an increase of .11% per year (95% CI, .12%–.13%) for the adjusted analysis, but these point estimates still reached statistical significance.

We would like to emphasize that this study represents a proof of principle and that the details of the method should ideally be replicated and calibrated for each sequencing laboratory or method (especially in the light of our finding of significant differences between laboratories) (Table 1). Moreover, it is important to note that 10%–20% of recently infected patients do have a high viral diversity because they are infected with several strains. Therefore, the fraction of ambiguous nucleotides should be only one of several measurements used to decide whether a patient is recently infected.

Detection of ambiguous nucleotides is a byproduct of bulk sequencing. Here, we have shown that this byproduct carries important information on the age of the infection at sampling time. In particular, a large frequency of ambiguous nucleotides argues against a recent infection event. The qualitative pattern of an initially linearly increasing amount of diversity and subsequent saturation is consistent with the diversification pattern observed for the env gene [3]. Overall, our study highlights the usefulness of diversity measures as markers for infection age. This usefulness might further increase with the advent of array-based pyrosequencing, which allows the diversity of the HIV population to be analyzed in much more detail.

### Supplementary Material

Supplementary materials are available at Clinical Infectious Diseases online (http://www.oxfordjournals.org/our_journals/cid/).

Supplementary materials consist of data provided by the author that are published to benefit the reader. The posted materials are not copyedited. The contents of all supplementary data are the sole responsibility of the authors. Questions or messages regarding errors should be addressed to the author.

### Acknowledgments

We thank the patients participating in the SHCS for their commitment, all the study nurses and study physicians for their invaluable work, the data center for data management, all the resistance testing laboratories for their high-quality work, and SmartGene for providing an impeccable database service. Furthermore, we thank Joseph Wong for critical reading of the manuscript.

**Financial support.** This work was supported by the Swiss National Science Foundation (grant number 3345-062041 in the framework of the SHCS, grant number 3247B0-112594 to H.F.G., S.Y., B.L., and S.B., grant...
Potential conflicts of interest. H.F.G. has been an adviser and/or consultant for GlaxoSmithKline (GSK), Abbott, Novartis, Boehringer Ingelheim, Roche, Tibotec, and Bristol-Myers Squibb (BMS) and has received unrestricted research and educational grants from Roche, Abbott, BMS, GSK, Tibotec, and Merck Sharp & Dohme (MSD). S.Y. has participated in advisory boards of BMS and Tibotec and has received travel grants from GSK and MSD. M.C. has received travel grants from Abbott, Boehringer Ingelheim, and Gilead. E.B. has been an advisor and/or consultant for Gilead and Abbott, has been a member of an advisory board of ViiV, Gilead, Tibotec, Pfitzer, and MSD, and has received research grants from Gilead and Abbott as well as travel grants from BMS, Gilead, ViIV, MSD, Abbott, and Tibotec. P.L.V. has been a member of an advisory board of MSD, Tibotec, Gilead, and ViIV and has received payment for lectures from Gilead, Tibotec, and GSK. All other authors report no potential conflicts.

References
1. Keele BF, Giorgi EE, Salazar-Gonzalez JF, et al. Identification and characterisation of transmitted and early founder virus envelopes in primary HIV-1 infection. Proc Natl Acad Sci USA 2008; 105:7552–7.
2. Bonhoeffer S, Holmes EC, Nowak MA. Causes of HIV diversity. Nature 1995; 376:125.
3. Shankarappa R, Margolick JB, Gange SJ, et al. Consistent viral evolutionary changes associated with the progression of human immunodeficiency virus type 1 infection. J Virol 1999; 73:10489–502.
4. Kahn JO, Walker BD. Acute human immunodeficiency virus type 1 infection. N Engl J Med 1998; 339:33–9.
5. Kouyos RD, von Wyl V, Yerly S, et al. Molecular epidemiology reveals long-term changes in HIV type 1 subtype B transmission in Switzerland. J Infect Dis 2010; 201:1488–97.
6. Schoeni-Affolter F, Ledergerber B, Rickenbach M, et al. Cohort profile: the Swiss HIV Cohort Study. Int J Epidemiol 2010; 39:1179–89.
7. Ledergerber B, Egger M, Opravil M, et al. Clinical progression and virological failure on highly active antiretroviral therapy in HIV-1 patients: a prospective cohort study. Lancet 1999; 353:863–8.
8. Von Wyl V, Yerly S, Boni J, 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.
9. Yerly S, Vora S, Rizzardi P, et al. Acute HIV infection: impact on the spread of HIV and transmission of drug resistance. AIDS 2003; 15:2287–92.
10. Boni J, Pyra H, Gebhardt M, et al. High frequency of non-B subtypes in newly diagnosed HIV-1 infections in Switzerland. J Acquir Immune Defic Syndr 1999; 22:174–9.
11. Yerly S, von Wyl V, Ledergerber B, et al. Transmission of HIV-1 drug resistance in Switzerland: a 10-year molecular epidemiology survey. AIDS 2007; 21:2223–9.
12. Joos B, Fischer M, Kuster H, et al. HIV rebounds from latently infected cells, rather than from continuing low-level replication. Proc Natl Acad Sci U S A 2008; 105:16725–30.
13. Taffe P, May M. A joint back calculation model for the imputation of the date of HIV infection in a prevalent cohort. Stat Med 2008; 27:4835–53.
14. Rieder P, Joos B, von Wyl V, et al. HIV-1 transmission after cessation of early antiretroviral therapy among men having sex with men. AIDS; 24:1177–83.
15. Wright S. Evolution in Mendelian populations. Genetics 1931; 16:0097–159.
16. Kouyos RD, Althaus CL, Bonhoeffer S. Stochastic or deterministic: what is the effective population size of HIV-1? Trends Microbiol 2006; 14:507–11.