Published in final edited form as:

Infect Genet Evol. 2008 September ; 8(5): 545–552. doi:10.1016/j.meegid.2008.03.004.

Multiple HIV-1 introductions into the Swedish intravenous drug user population

Helena Skar^{1,2,3}, Staffan Sylvan⁴, Hans-Bertil Hansson⁵, Olle Gustavsson^{6,7}, Hans Boman⁷, Jan Albert^{1,2}, and Thomas Leitner^{3*}

¹Department of Virology, Swedish Institute for Infectious Disease Control, SE-17182 Solna, Sweden.

²Dept of Microbiology Tumor and Cell Biology, Karolinska Institutet, SE-17177, Stockholm, Sweden.

³Theoretical Biology and Biophysics, Group T-10, MS K710, Los Alamos National Laboratory, Los Alamos, NM 87545, USA

⁴Regional Center for Infectious Disease Control & Prevention, Dag Hammarskjölds väg 17, SE-751 85 Uppsala, Sweden

⁵Regional Center for Infectious Disease Control & Prevention, University Hospital MAS, SE-20502, Malmö, Sweden

⁶Department of Clinical Microbiology, University Hospital, SE-58185 Linköping, Sweden

⁷Department of Communicable Diseases Control, Sundsvall Hospital, SE-851 86 Sundsvall, Sweden

Abstract

In 2001, an increase of HIV-1 diagnoses among intravenous drug users (IVDU) was reported in Sweden. In nearby countries, Finland, Russia and the Baltic states, recent outbreaks had been described with thousands of new cases over short periods of time. Since there was a concern that these outbreaks would carry over to Sweden a study was initiated to determine the factors leading to the Swedish increase of HIV-1 diagnosed IVDUs. HIV-1 *env* V3 sequences were obtained from 97 patients enrolled in ongoing epidemiological studies encompassing the years 1987–2004 with a focus on 2001–2002. The sequences were used for maximum likelihood and Bayesian inference of the molecular epidemiology. Among the virus spreading in 2001–2002, we found that four different subtypes/CRFs were present in the Swedish IVDU population (A, B, CRF01_AE and CRF06_cpx). Subtype B constituted 85% of the infections, established by 12 independent introductions into the IVDU population. The worrisome increase in 2001 was mainly not a result of import of the outbreaks in nearby countries, but rather a higher detection rate of secondary cases due to efficient epidemiological tracing of the generally slow spread of established forms of subtype B in the IVDU community. However, a few of the non-subtype B cases were linked to the outbreaks in Finland, Estonia and Latvia. Because HIV-1 outbreaks easily can be exported from one country to another amongst IVDUs, this prompts continued surveillance in the Baltic Sea Region.

*Corresponding author, email: tkl@lanl.gov.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Keywords

HIV-1; molecular epidemiology; outbreak; evolutionary rate; phylogeny; tree node height; intravenous drug users

INTRODUCTION

AIDS was first diagnosed in Swedish patients in 1982–1983 (Pehrson P.O., 1983). In the early phase of the Swedish epidemic 27% of the diagnosed patients were intravenous drug users (IVDUs) (Heinrup, 2005). The spread within this risk group was first reported in the years of 1984–85 in the region of Stockholm, but was encouragingly followed by a drastic decrease in newly diagnosed cases in the following years. Other regions were not as affected with HIV outbreaks among their IVDU populations at this time and this trend has sustained. Only sporadic cases and sometimes a few secondary cases have been reported from other cities such as Gothenburg, Malmö and Sundsvall. The number of newly diagnosed HIV infections among IVDUs stabilized to around 20 cases per year in the nineties, but in 2001 there was a sharp increase (38 in contrast to an average of 19 per year during the preceding five years). At approximately the same time there were large outbreaks of HIV-1 infections among IVDUs in countries of the former Soviet Union and also a smaller outbreak in Finland (Liitsola et al., 2000). A concern was raised that the increase in newly diagnosed HIV infections among Swedish IVDUs in 2001 could be linked to East European outbreaks.

HIV-1 is divided into three genetic groups (M, N, O) and group M is further divided into nine genetic subtypes A through K as well as many circulating recombinant forms (CRFs). Subtype B dominates completely among IVDUs in Western Europe and the U.S. (Aceijas et al., 2004; Lukashov et al., 1996; Op de Coul et al., 2001; Robbins et al., 2003), whereas subtype A has been shown to dominate in Eastern Europe and the regions of the former Soviet Union (Bobkov A et al., 1997; Bobkov et al., 1998; Ferdats et al., 1999). Some CRFs have also been reported to cause local outbreaks among IVDUs in Europe; CRF03_AB in Russia (Liitsola et al., 1998), CRF01_AE in Finland (Kivela et al., 2005; Liitsola et al., 2000a; Liitsola et al., 2000b), and more recently, CRF06_cxp in Estonia (Adojaan et al., 2005; Zetterberg et al., 2004).

The aim of this study was to map the molecular epidemiology of HIV-1 among IVDUs in Sweden in order to understand the underlying mechanisms of the sharp increase of newly HIV infected IVDUs in 2001. It was feared that the increase was due to introduction of HIV-1 from the former Soviet Union and that the increase was the start of a larger outbreak also in Sweden. By detailed molecular epidemiology analysis we could show that a majority of the new cases were infected with “traditional” Western European subtype B variants and that the contribution from the East European HIV-1 variants was small.

MATERIALS & METHODS

Study population

Blood samples (EDTA treated) were collected as part of ongoing epidemiological studies from 47 HIV-1-infected individuals belonging to the IVDU risk group between the years of 2001 and 2002, which represent 66% of all the diagnosed IVDU cases in Sweden during these two years. In general, the vast majority of HIV-1 cases among IVDUs in Sweden are diagnosed in the Stockholm region but during the increase of IVDU cases in 2001–2002 a local outbreak in Sundsvall was also detected. Thus, the patients diagnosed in the Stockholm and Sundsvall region were included in this study (Figure 1). Among these, 34 were male, 10 were female and for 3 the gender was unknown.

We also included 50 local controls consisting of Swedish IVDUs diagnosed in 1987–2004, except 2001 and 2002. These sequences came from IVDUs who had been HIV-1 diagnosed in the regions of Stockholm, Uppsala, Sundsvall or Malmö.

The samples were collected, transported and analyzed under code, thus the identity of the persons participating in this study were unknown for the researchers.

This study was approved by the Regional Research Ethics Committee in Stockholm, Dnr: 2005/944-31/1.

DNA Sequencing

Plasma was separated from EDTA-treated whole blood through centrifugation at 2500 rpm, 10 minutes and then stored at -70°C . Viral RNA was extracted from plasma using the Nuclisense RNA extraction kit (NASBA Diagnostics, Organon Teknika, Boxtel, The Netherlands) according to the manufacturer's instruction, and reverse transcribed with First-Strand cDNA Synthesis Kit (Amersham Pharmacia Biotech, Uppsala, Sweden). Nested PCR was carried out amplifying a 410-base pair fragment of the HIV *env* gene corresponding to the V3 domain and flanking regions (Leitner et al., 1996b). The PCR fragments were subsequently purified with GFX PCR DNA and Gel Band Purification Kit (Amersham Biosciences, Uppsala, Sweden). Cycle sequencing was performed on both strands using the Big Dye Terminator Cycle Sequencing Ready Reaction Kit (Applied Biosystem, Foster City, CA, US). Prior to detection, the DNA fragments were purified with a dye terminator removal and sequencing reaction clean-up process using Sephadex G-50 Superfine (Amersham Pharmacia Biotech, Uppsala, Sweden) in a Multiscreen 96-well filtration plate (Millipore Corporation). The fragments were detected on an ABI PrismTM 3100 DNA capillary sequencer and the nucleotide sequences were examined using the software program SequencerTM (Gene Codes Corporation, Ann Arbor, MI, US). The sequences are deposited in Genbank under the accession numbers EU010264–EU010360.

Phylogenetic inference

Sequences were manually aligned in the correct reading frame using the program Se-Al (Rambaut, 1996) to previously published relevant sequences, including the 2005 HIV subtype reference sequences from the HIV database (Leitner et al., 2005), three sequences from the Estonian outbreak of the circulating recombinant form CRF06_cpx among IVDUs (Adojaan et al., 2005; Zetterberg et al., 2004), and other closely related sequences as detected by BLAST and HIV database searches on relevant geographic region, risk groups, subtype and genomic coordinates (www.hiv.lanl.gov). Phylogenetic analyses were performed using maximum likelihood analysis on a parallel computing cluster. The software was an optimized and expanded parallel code based on Olsen's fastDNAm1 (Olsen et al, 1994) and DNArates (DNArates was written by Gary Olsen and colleagues at the Ribosomal Database Project (RDP) at the University of Illinois at Urbana-Champaign), which includes a general-time-reversible model with unequal rates among nucleotide sites. Bootstrap analysis was performed using 1000 replicates with the neighbour-joining method under the F84 model with the program PAUP* (Swofford DL, 2002).

Estimating the number of introductions

Local sub-epidemics and imported unlinked cases in the Swedish IVDU subtype B and CRF01_AE populations were investigated using node height distances in rooted maximum likelihood (ML) trees. For each relevant subtype a universal reference dataset was created consisting of epidemiologically “unlinked” sequences from the HIV database so that all sequences matched the sampling years of the query dataset and multiple samples from the same patients were excluded. Based on ML trees a least square rooting algorithm was implemented

(Athreya et al., unpublished data; Maljkovic Berry et al., 2007). The distances from the root to each node in the tree were calculated and these node heights form a distribution that displays the number of lineages at a certain distance from the most recent common ancestor of the subtype in question. For the universal reference dataset, the rooting point reflects the first divergence event of that subtype, *i.e.*, the cenancestor. This is not always the case for the query dataset why a calibration was needed. Therefore, the average tree height of the reference dataset and the query dataset were calculated, respectively. Since we matched the reference dataset and query dataset according to time of sampling, the average tip height of the two trees could be aligned, giving us a node height distribution for the query dataset displaying the distance from the most recent common ancestor of the subtype in question. To attempt to make a less subjective call for what “unlinked” means, we calculated a cut-off value for “unlinked” simply by the upper 99% percentile of the reference sequence distribution. Finally, we compared the node height distribution of the query sequence population to the calculated cut-off. Consequently, this will give us an estimate of how many “unlinked” introductions, as defined by the universal reference distribution, we have in the query population. We evaluated the robustness of the cut-off formed by the upper 99% percentile of the reference node height distribution by conducting 100 bootstrap replicates. The mean of the bootstrap distribution along with the 95% confidence interval was used as cut-off value. Recall that the included sequences are what typically are considered epidemiologically “unlinked” reference sequences from the HIV database. They are derived from different publications, different authors and different laboratories.

We validated our method on two previously published datasets: One set of 13 Russian subtype B sequences sampled between the years 1991 and 1995 (Leitner et al., 1996b), and one set of 38 Estonian subtype B sequences sampled between the years 1995 and 1999 (Ustina et al., 2001). We searched in the HIV database for subtype B sequences sampled between these years and also conducted BLAST searches for each of the strains. The universal reference subtype B dataset was modified to match the sampling years of the query datasets. Confirmatory phylogenetic trees were made using BioNJ under the F84 substitution model in PAUP* (Swofford DL, 2002).

Timing of introduction events

After identifying independent Swedish transmission chains among the subtype B sequences, we removed all global reference sequences, saving only Swedish sequences belonging to a larger local transmission chain. Coalescent-based analysis was used to estimate the time of the most recent common ancestor (T_{MRCA}) for each of the clusters, by means of a Bayesian MCMC (Markov Chain Monte Carlo) method as implemented in the BEAST program (<http://beast.bio.ed.ac.uk/>) (Drummond et al., 2006). The Bayesian Skyline demographic model with 3–6 steps was used assuming either a strict or a relaxed molecular clock. The Bayesian Skyline demographic model was used since this population model makes the least assumptions about the actual population history and we were only interested in estimating substitution parameters and T_{MRCA} for certain clusters. Because each such cluster did not contain enough information by itself to accurately estimate the evolutionary model parameters, we combined all transmission chains described by an independent tree but sharing the same substitution parameters into a single MCMC run. To investigate if the evolutionary rate for the three transmission chains were the same, individual MCMC runs for each transmission chain was performed. The general time-reversible model of nucleotide substitution was employed with a gamma distribution for among-site rate variation. Bayes Factor of the estimated marginal likelihood in respect to the prior was used to discriminate between the different clock models to determine which model best fitted our data.

RESULTS

Limited Import of HIV-1 from Eastern Europe

This study included 47 IVDUs diagnosed in Sweden in 2001 and 2002, which formed our cohort, and another 50 HIV-1 local control sequences drawn from IVDUs in other years between 1987 and 2004. According to our phylogenetic analysis four different subtypes or recombinant forms were present in the Swedish IVDU cohort. The majority ($n=40$, $>85\%$) of the sequences was of subtype B, and two patients carried subtype A, four patients carried CRF01_AE and one patient carried CRF06_cpx (Figure 2).

The two subtype A sequences were compared to other subtype A sequences through BLAST and HIV database searches and phylogenetic analyses. Patient 57 carried a subtype A variant that was closely related to subtype A Latvian sequences (data not shown), while patient 70 carried a sequence that was related to African subtype A sequences (data not shown). Patient 64, who originated from Russia, carried a CRF06_cpx variant, which was closely related to viruses in the recent outbreak of CRF06_cpx infections among Estonian IVDUs (Adojaan et al., 2005; Zetterberg et al., 2004), see Figure 2.

As described in more detail below, but evidenced already by the subtype distribution, the analysis strongly indicate that the increase of HIV-1 diagnoses among Swedish IVDUs in 2001 was not due to a large import of HIV variants from Eastern Europe.

Mostly Local Spread of Subtype B Variants

The Swedish subtype B IVDU sequences were further analyzed together with previously published global reference sequences belonging to different risk groups as well as local control sequences from Swedish IVDUs diagnosed before 2001 or after 2002 (Figure 3). The phylogenetic tree revealed that three larger transmission chains (clusters) were present in our cohort. The clusters were named according to their geographic localization: Stockholm clusters I and II, and the Sundsvall cluster. Some interactions of the geographical regions were however observed: one Sundsvall sequence appeared in Stockholm cluster I, and one Uppsala sequence appeared in the Sundsvall cluster.

The global reference sequences that intermixed with our Swedish sequences were mostly sampled in neighboring European countries, but also included a few sequences from Switzerland, the Netherlands, the U.S.A. and Asian countries. When interpreting the tree a lineage was assumed to be independent if it was not split by foreign reference sequences. Strictly following that rule 33 independent lineages would be inferred among all Swedish IVDU sequences. Among sequences collected during 2001–2002 there would be 13. Furthermore, in two cases only one reference sequence (Estonian reference sequences AF307255 and AF307254, respectively) split a Swedish lineage. Hence, these Swedish lineages could either be regarded as independent (*i.e.*, 13 independent lineages) or alternatively the Estonian sequences could be regarded to belong to the Swedish lineages (*i.e.*, 11 independent lineages).

Multiple introductions of subtype B

As described above, according to the ML tree in Figure 3, the Swedish 2001–2002 subtype B dataset appeared to have 11–13 independent lineages, which were split by global, non-Swedish, sequences. In an attempt to further assess whether these independent lineages reflected independent introductions of HIV-1 into the Swedish IVDU population, we investigated the node height distribution in a phylogenetic tree (Figure 4). A universal reference dataset, matched to subtype, sampling year and filtered to include only one sequence per patient, were used to estimate a node height that described what is commonly considered as a level at where

sequences are epidemiologically unlinked. The reference dataset indicated that at 0.091 ± 0.001 substitutions/site (derived from the upper 99% percentile of the reference node height distribution) from node 0 (the root of the tree) we would expect 12 epidemiologically unlinked lineages in the Swedish IVDU population. As seen in Figure 4 and in Table 1, the node height distribution agreed well with the ML tree estimate. For instance, the split between the Stockholm II and Sundsvall clusters (and the sequences between them) has a node height that is not statistically different from that of random reference sequences from the HIV database (matched to subtype and sampling year, and filtered to include only one sequence per patient). Thus, the linkage between these clusters goes back in time to a point where sequence associations are generally considered “unlinked” and that probably predates the introduction of HIV into the Swedish IVDU population.

Validation of method used to infer number of introductions

The method to estimate the number of introductions was validated using two previously published datasets consisting of Russian and Estonian sequences. In Leitner et al 1996b, subtype B had entered Russia on at least four occasions, and in addition there had been HIV-1 exchange between Lithuanian and Russian homosexual men. With our method we estimated four introductions into the examined Russian population. According to Ustina et al 2001, there were several introductions of HIV-1 subtype B into the Estonian population they studied; their phylogenetic tree suggested at least six independent lineages, and our method estimated seven unlinked introductions (Table 1). Thus, the estimates of the number of introductions agreed well between the original phylogenetic analyses and our new method.

Timing the introductions of subtype B clusters

A Bayesian analysis was carried out to determine the ages of Stockholm clusters I and II and the Sundsvall cluster. Three different models of molecular clock evolution (strict-, relaxed uncorrelated lognormal-, and relaxed uncorrelated exponential-clock) were tested and they fitted the data equally well according to the Bayes Factor. The estimates of the substitution parameters along with T_{MRCA} for each cluster are presented in Table 2. The mean substitution rate was estimated to be $7.2\text{--}10.7 \times 10^{-3}$, which is compatible to previously published estimates derived from partial HIV-1 *env* sequences, [2.4×10^{-3} (Korber et al., 2000), 4.7×10^{-3} (Robbins et al., 2003), 6.7×10^{-3} (Leitner et al., 1996a), and $2.02\text{--}16.9 \times 10^{-3}$ (Maljkovic Berry et al., 2007)]. The age of the three subtype B clusters present in the Swedish IVDU population differed. The Stockholm clusters I and II were estimated to have entered the population in the mid 1990s, with T_{MRCA} dated to 1995 and 1997, respectively. The Sundsvall cluster seemed to be a newer introduction with an estimated time of entry dated to 1999.

In addition, we conducted separate BEAST analyses for each transmission chain to check if the clusters had similar evolutionary rates. Only Stockholm cluster I gave meaningful results, whereas the Stockholm cluster II and the Sundsvall cluster had too few taxa. The substitution rate for Stockholm cluster I was approximately 33% lower for all three molecular clock models when run individually, and T_{MRCA} was calculated to be 1.5 years older (1993) than in the combined analysis.

CRF01_AE in Swedish IVDUs

Four patients infected with CRF01_AE were found (Figure 5). In a preliminary analysis we observed that one Finnish reference sequence, 01_AE.FI.93.Ht.AF219378, grouped together with three of the Swedish sequences. We therefore contacted Dr. Mika Salminen in Finland who kindly provided CRF01_AE sequences from the Finnish IVDU outbreak, (Kivela et al., 2005). Figure 5 shows that the Finnish sequences formed a tight cluster, and that the three Swedish sequences group together with these sequences, even though they do not intermix. Patient 39 had epidemiological connections to Finland, which could explain an introduction

of this variant of CRF01_AE into the Swedish IVDU population. Patient 82 is probably a single independent introduction since the sequence from this patient mix with sequences from IVDUs in Thailand.

DISCUSSION

We have investigated the molecular epidemiology of HIV-1 infected IVDUs in Sweden in order to understand if the increase of HIV-1 diagnoses among Swedish IVDUs in 2001 was due to increasing diagnoses of domestic transmissions or import from abroad. The main finding was that the vast majority of the cases diagnosed in 2001–2002 were infected with local subtype B variants, and that a few infections represented influx of East European HIV-1 variants.

The phylogenetic analysis showed that four different subtypes and CRFs were present in our study cohort. Only a total of seven patients in our cohort carried non-B subtypes. In five cases variants were present in Eastern European outbreaks: subtype A in Latvia, CRF01_AE in Finland and CRF06_cpx in Estonia. Our finding that subtype B constituted the majority of our sequences agrees with earlier studies made on IVDUs in Western Europe (Hamers et al., 1997; Lukashov et al., 1996; Op de Coul et al., 2001).

Among the patients infected with the Finnish variant of CRF01_AE, one had an epidemiological connection to Finland which may explain the introduction of this variant in Sweden. Patient 64, which was the only patient within our cohort that carried CRF06_cpx, originated from Russia. The geographical proximity of this patient's origin plausibly explains how this CRF had entered Sweden from the massive outbreak among IVDUs in Estonia. Among the patients carrying subtype B, regional clustering of the sampled sequences could be seen. The sequences sampled from patients living in Sundsvall formed one distinctive transmission chain, while the sequences sampled in Stockholm formed two distinctive transmission chains, although some overlap was noted. This indicates that most transmissions are local, with sporadic interactions between social networks from different parts of Sweden. The Sundsvall cluster has been carefully followed up, and the epidemiological information on these patients correlates well with the information obtained by the phylogenetic tree and the Bayesian analysis (Figure 3, Table 2). The T_{MRCA} for the Sundsvall cluster was estimated to be in mid to late 1999, and patient 75 had his last negative HIV test in August 1998 while patients 31, 53 and 54 had their last negative HIV tests in February, May and June 2000, respectively. Among the sequences from 2001–2002, the ML tree analysis also suggested that 5 single subtype B sequences represent patients who independently may have attracted their infections abroad. The epidemiological data on one of these patients agrees with this, *i.e.*, patient 59 originates from the Netherlands and has stated a long-term use of intravenous drugs. Concurring, the sequence from this patient (P59.EU010329) clustered with sequences from the Netherlands drawn from IVDUs. Patient 43 originated from Sweden and was diagnosed with HIV as well as AIDS in 2001. The sequences closest to sequence P43.EU010314 were obtained from Swiss IVDUs. Patients 44, 56 and 62 originated from Sweden but we don't have any more data about these patients.

The sharp increase in 2001 of HIV-1 among IVDUs was mostly the consequence of spread within Sweden of already established HIV-1 forms rather than rampant import from nearby outbreaks. Whether the augmented number of newly diagnosed cases in 2001 was a result of a direct increase of newly HIV-1 infected IVDUs or of an enhanced surveillance resulting in more diagnoses during this time period remains somewhat unclear. While the number of general HIV tests probably has been fairly constant over the past 10 years in the Stockholm region, in Sundsvall the number of HIV tests had actually declined in 2001. However, when a new HIV case is discovered, Swedish law states that the source of the infection and possible secondary infections must be traced and patients informed. Therefore, if there are any

secondary cases, they are usually discovered fairly soon after the index case, explaining why the number of general HIV tests may not correlate with the number of newly diagnosed cases. After the year of 2001 the number of newly diagnosed IVDUs in Sweden slowly decreased, supporting the hypothesis that the peak in 2001 was due to an increase in follow-up diagnoses rather than an increased spread of HIV-1. On the other hand, the CRF01_AE infections in Sweden are very recent, as indicated by the short branches between sequences from three of the patients in Figure 5. Recall that these three sequences were closely related to the Finnish outbreak.

The phylogenetic analysis of the Swedish subtype B dataset suggested that there existed 11–13 lineages among the sequences that were spreading in 2001–2002. Traditional tree analysis is dependent on the existence of “good” reference sequences that can split query lineages and thereby indicate separate introductions. Therefore, this type of tree-based analysis may underestimate the true number of introductions into a population if good references are lacking. To address this problem we attempted to develop an alternative method to discriminate between linked and unlinked sequences. Thus, we used a universal reference dataset to calculate the average distance between “unlinked” sequences in a tree. The method is meant to work as a rough estimation tool when good reference sequences are lacking. The method formalizes the process of retrieving reference sequences and makes a less subjective interpretation of how many introductions have occurred into a study population. The output is an estimate of the number of lineages in a tree at a certain distance. When the method was tested on our subtype B dataset the estimated number of lineages agreed well with the number estimated by a traditional phylogenetic tree analysis (12 versus 11–13). This was expected because we had access to a large set of good reference sequences for the phylogenetic tree analysis. The validation datasets from Russia and Estonia showed also good agreement with the published traditional tree analyses (Leitner et al., 1996b; Ustina et al., 2001).

It is important to note that the node height method is dependent upon the non-linkage of the sequences in the reference dataset, as is the traditional tree based interpretation. We believe that our selection procedure of the reference dataset is adequate since we include all available sequences of the same subtype (and genomic region) in the HIV database, excluding only multiple sequences from the same patient and multiple sequences sampled from known sub-epidemics and transmission chains. Although the method gave concordant results when applied to HIV-1 subtype B it should be noticed that the method was not applicable to CRF01_AE, since it was not possible to retrieve a sufficiently large reference dataset due to a lack of sequences from different geographical regions during the corresponding years from the HIV database.

When the results from the Bayesian analysis are put into context of the ML tree in Figure 3, one could argue that the two analyses gave discordant results. Looking at the genetic distance from the root in the ML tree, Stockholm cluster II seems to be a newer introduction than the Sundsvall cluster. In the Bayesian analysis the time of entry is reversed for the two clusters. This outcome can have several potential explanations. First of all, the ML tree is a non-molecular clock tree, which means that the evolutionary rate is free to vary over all branches. As can be seen in the ML tree in Figure 3, the taxa are not necessarily placed correctly on a time-scale in regards to time of sampling. The BEAST analysis enforces a molecular clock (in our case; strict, relaxed uncorrelated log-normal, and relaxed uncorrelated exponential clock) and takes time of sampling into account. This means that the molecular-clock corrected tree may not explain the data best, because the clock has to be enforced to be able to calculate T_{MRCA} for the three clusters of interest. Furthermore, in our BEAST analysis we let three independent trees, *i.e.*, the three transmission clusters, run in the same MCMC run while sharing the same substitution parameters and rate parameter in the case of the strict clock, and with the rate parameters being drawn from the same distribution in the cases of the relaxed clocks. This

means that we postulate that the three clusters actually have the same evolutionary rate. Unfortunately the number of taxa in two of the three clusters was too few to determine if this was the case. It has recently been shown that the evolutionary rate may differ due to differences in epidemiological scenarios (*i.e.*, fast spread vs. slow spread) (Maljkovic Berry et al., 2007). This may apply even though all our samples were drawn from IVDUs. Maljkovic Berry et al. also noticed that as samples were closer in time to each other the stochastic error became greater, *i.e.*, the estimated rate variation increased with shorter time span between samples. This appeared to be especially the case for subtype B, where the time needed to give a stable estimation of evolutionary rate on the population level was approximately 10 years. Because the Stockholm cluster II and the Sundsvall cluster each had a time span of about 4 years, this may have influenced the accuracy in the analyses. Interestingly, our estimation of the evolutionary rate of subtype B present in the local epidemics among IVDUs in Sweden in 2001–2002 is high ($7.2\text{--}10.7 \times 10^{-3}$ subs site⁻¹ yr⁻¹) which, according to Maljkovic Berry et al. (2007), is evidence for a slow spreading epidemic.

In conclusion, we have shown that the Swedish HIV-1 IVDU epidemic has had many introductions over time and is currently composed of 4 subtypes/CRFs. The worrisome increase in 2001 was not a result of import of the outbreaks in nearby countries, but rather a continuation of the generally slow spread of established forms of subtype B in the IVDU community. However, in late 2006 a new outbreak of HIV-1 has been reported among the IVDUs in Stockholm. Thus, the initial concern of import from the massive outbreaks in Eastern Europe may once again be valid, prompting continued surveillance.

ACKNOWLEDGEMENTS

We thank Marcus G. Daniels, Kajsa Aperia, Afsaneh Heidarian and Maj Westman for excellent technical assistance. We would also like to thank Alexei Drummond for helpful tips regarding the BEAST analyses, and Mika Salminen who kindly provided us with sequences from an outbreak of HIV-1 among IVDUs in Finland. This study was funded by an NIH-DOE interagency agreement (Y1-A1-1500-06), funds from The Swedish Research Council and the Swedish National Institute of Public Health.

REFERENCES

- Aceijas C, Stimson GV, Hickman M, Rhodes T. Global overview of injecting drug use and HIV infection among injecting drug users. *AIDS* 2004;18(17):2295–2303. [PubMed: 15577542]
- Adojaan M, Kivisild T, Männik A, Krispin T, Ustina V, Zilmer K, Liebert E, Jaroslavtsev N, Priimägi L, Tefanova V, Schmidt J, Krohn K, Villems R, Salminen M, Ustav M. Predominance of a Rare Type of HIV-1 in Estonia. *J. Acquir. Immune Defic. Syndr* 2005;39(5):598–605. [PubMed: 16044014]
- Bobkov A, Cheingsong-Popov R, Selimova L, Ladnaya W, Kazennova E, Krawchenko A, Fedotov A, Saukhat S, Zverev S, Pokrovsky V, Weber J. An HIV type 1 epidemic among injecting drug users in the former Soviet Union caused by a homogenous subtype A strain. *AIDS Res. Hum. Retroviruses* 1997;13:1195–1201. [PubMed: 9310286]
- Bobkov A, Kazennova E, Selimova L, Bobkova M, Khanina T, Ladnaya N, Kravchenko A, Pokrovsky V, Cheingsong-Popov R, Weber J. A sudden epidemic of HIV type 1 among injecting drug users in the former Soviet Union: identification of subtype A, subtype B, and novel gagA/envB recombinants. *AIDS Res. Hum. Retroviruses* 1998;14(8):669–676. [PubMed: 9618078]
- Drummond AJ, Ho SY, Phillips MJ, Rambaut A. Relaxed phylogenetics and dating with confidence. *PLoS Biol* 2006;4(5):e88. [PubMed: 16683862]
- Ferdats A, Konicheva V, Dievberna I, Lilja E, Albert J. An HIV Type 1 Subtype A Outbreak among Injecting Drug Users in Latvia. *AIDS Res. Hum. Retroviruses* 1999;15(16):1487–1490. [PubMed: 10555112]
- Hamers FF, Batter V, Downs AM, Alix J, Cazein F, Brunet JB. The HIV epidemic associated with injecting drug use in Europe: geographic and time trends. *AIDS* 1997;11(11):1365–1374. [PubMed: 9302447]

- Heinrup, M. HIV epidemin i Sverige, Vol. SMI-rapport 2:2005. Smittskyddsinstitutet. Sweden: Solna; 2005. (In Swedish)
- Kivela PS, Krol A, Salminen MO, Gekus RB, Suni JI, Anttila VJ, Liitsola K, Zetterberg V, Miedema F, Berkhout B, Coutinho RA, Ristola MA. High plasma HIV load in the CRF01-AE outbreak among injecting drug users in Finland. *Scand. J. Infect. Dis* 2005;37(4):276–283. [PubMed: 15871166]
- Korber B, Muldoon M, Theiler J, Gao F, Gupta R, Lapedes A, Hahn BH, Wolinsky S, Bhattacharya T. Timing the ancestor of the HIV-1 pandemic strains. *Science* 2000;288(5472):1789–1796. [PubMed: 10846155]
- Leitner T, Escanilla D, Franzen C, Uhlen M, Albert J. Accurate reconstruction of a known HIV-1 transmission history by phylogenetic tree analysis. *Proc. Natl. Acad. Sci. U.S.A* 1996a;93(20):10864–10869. [PubMed: 8855273]
- Leitner, T.; Foley, B.; Hahn, B.; Marx, P.; McCutchan, F.; Mellors, J.; Wolinsky, S.; Korber, B. HIV Sequence Compendium 2005. New Mexico, U.S.A.: Published by Theoretical Biology and Biophysics Group, Los Alamos National Laboratory; 2005.
- Leitner T, Korovina G, Marquina S, Smolskaya T, Albert J. Molecular epidemiology and MT-2 cell tropism of Russian HIV type 1 variant. *AIDS Res. Hum. Retroviruses* 1996b;12(17):1595–1603. [PubMed: 8947294]
- Liitsola K, Holmström P, Laukkanen T, Brummer-Korvenkontio H, Leinikki P, Salminen M. Analysis of HIV-1 Genetic Subtypes in Finland Reveals Good Correlation Between Molecular and Epidemiological Data. *Scand. J. Infect. Dis* 2000a;32(5):475–480. [PubMed: 11055649]
- Liitsola K, Ristola M, Holmstrom P, Salminen M, Brummer-Korvenkontio H, Simola S, Suni J, Leinikki P. An outbreak of the circulating recombinant form AECM240 HIV-1 in the Finnish injection drug user population. *AIDS* 2000b;14(16):2613–2615. [PubMed: 11101077]
- Liitsola K, Tashkinova I, Laukkanen T, Korovina G, Smolskaja T, Momot O, Mashkilleyson N, Chaplinskas S, Brummer-Korvenkontio H, Vanhatalo J, Leinikki P, Salminen M. HIV-1 genetic subtype A/B recombinant strain causing an explosive epidemic in injecting drug users in Kaliningrad. *AIDS* 1998;12:1907–1919. [PubMed: 9792392]
- Lukashov V, Kuiken C, Vlahov D, Coutinho R, Goudsmit J. Evidence for HIV type 1 strains of U.S. intravenous drug users as founders of AIDS epidemic among intravenous drug users in northern Europe. *AIDS Res. Hum. Retroviruses* 1996;12(12):1179–1183. [PubMed: 8844022]
- Maljkovic Berry I, Ribeiro R, Kothari M, Athreya G, Daniels M, Lee HY, Bruno W, Leitner T. Unequal evolutionary rates in the HIV-1 pandemic: The evolutionary rate of HIV-1 slows down when the epidemic rate increases. *J. Virol* 2007;81(19):10625–10635. [PubMed: 17634235]
- Op de Coul EL, Prins M, Cornelissen M, van der Schoot A, Boufassa F, Brettle RP, Hernandez-Aguado L, Schiffer V, McMenamin J, Rezza G, Robertson R, Zangerle R, Goudsmit J, Coutinho RA, Lukashov VV. Using phylogenetic analysis to trace HIV-1 migration among western European injecting drug users seroconverting from 1984 to 1997. *AIDS* 2001;15(2):257–266. [PubMed: 11216936]
- Pehrson PO, L K, Bergdahl S, Morfeldt-Månsson L, Blaxhult A, Broström C. AIDS bland homosexuella män: De första svenska fallen. *Läkartidningen* 1983;80:545–548. (In Swedish)
- Rambaut, A. Se-Al: Sequence Alignment Editor. 1996. <http://tree.bio.ed.ac.uk/software/seal/>
- Robbins KE, Lemey P, Pybus OG, Jaffe HW, Youngpairaj AS, Brown TM, Salemi M, Vandamme AM, Kalish ML. U.S. Human immunodeficiency virus type 1 epidemic: date of origin, population history, and characterization of early strains. *J Virol* 2003;77(11):6359–6366. [PubMed: 12743293]
- Swofford, DL. PAUP* Phylogenetic Analysis Using Parsimony (*and Other Methods) 4.0b10. Sunderland, MA: Ed. Sinauer Associates; 2002.
- Ustina V, Zilmer K, Tammai L, Raukas M, Andersson A, Lilja E, Albert J. Epidemiology of HIV in Estonia. *AIDS Res. Hum. Retroviruses* 2001;17(1):81–85. [PubMed: 11177386]
- Zetterberg V, Ustina V, Zilmer K, Liitsola K, Kalikova N, Sevastianova K, Brummer-Korvenkontio H, Leinikki P, Salminen M. Two Viral Strains and a Possible Novel Recombinant Are Responsible for the Explosive Injecting Drug Use-Associated HIV Type 1 Epidemic in Estonia. *AIDS Res. Hum. Retroviruses* 2004;20(11):1145–1147. [PubMed: 15588335]

Figure 1.

A map over Scandinavia and the Baltic countries where the Swedish cities of, Stockholm, Gothenburg, Malmö, Sundsvall and Uppsala are indicated.

Figure 2.

A maximum likelihood tree of Swedish and subtype reference *env* sequences retrieved from the Los Alamos HIV Database (Leitner et al., 2005). Swedish sequences sampled in the years of 2001 and 2002 are marked with a filled circle. Swedish local reference sequences are marked with an open circle. Three sequences from the local outbreak among IVDUs in Estonia are also included. Subtype B and CRF01_AE are here shown with a triangle since further analysis and more detailed trees will be shown below. Neither CRF01_AE nor CRF06_cpx has a breakpoint in the region analyzed here. * Indicates a bootstrap value > 70. Name code is as follows: Reference sequences: Subtype.Region of sampling.Year.Riskgroup.Accession number; Swedish sequences: Region of sampling.Year.Riskgroup.Patient number.Accession number.

Figure 3.

A detailed maximum likelihood tree of Swedish subtype B sequences and closely related sequences. Swedish IVDU sequences sampled in the years of 2001 and 2002 are marked with a filled circle. Swedish IVDU local reference sequences are marked with an open circle. Subtype D, (sequence ELI), was used as outgroup, and for clarity the outgroup has been cropped. A tree with all the taxon names can be retrieved from the authors upon request. * Indicates a relevant bootstrap value > 70.

Figure 4.

a) Universal reference dataset with epidemiologically unlinked subtype B sequences sampled during the years 2001 – 2002. Distance to node 0 is the inferred distance to the cen ancestor.
 b) Query dataset with Swedish subtype B sequences from IVDUs sampled during the years of 2001 – 2002. * The 99% upper percentile limit is shown, 100 replicates of bootstrap analysis was made to evaluate the robustness of this limit. The mean of the bootstrap distribution is used as cut-off value. ** Average tree height is calculated to enable calibration of the node heights from the query dataset.

Figure 5.

A detailed maximum likelihood tree of Swedish CRF01_AE sequences and closely related sequences. Swedish sequences sampled in the years of 2001 and 2002 are marked with a filled circle. Swedish local reference sequences are marked with an open circle. Subtype D, (sequence ELI), was used as outgroup, and for clarity the outgroup has been cropped. * Indicates a bootstrap value > 70.

Table 1
Summary of the results from the analyses based on node height distribution

Dataset	N _{Query}	Country of Sampling	Risk group	Time of sampling	Subtype	N _{Ref}	Cut-off ¹⁾ [subst. site ⁻¹]	I _{auth} ²⁾	I _{tree} ³⁾	I _{inh} ⁴⁾
This study	42	Sweden	IVDU	2001–2002	B	76	0.091 ± 0.001	N/A	11–13	12
Leitner et al. 1996b	13	Russia	MSM and Heterosexual	1991–1995	B	128	0.174 ± 0.003	≥4	4–11	4
Ustina et al. 2001	38	Estonia	MSM and Heterosexual	1995–1999	B	155	0.259 ± 0.003	≥6	7–12	7

1) Bootstrapped mean and 95% CI of cut-off distance for I_{inh}.

2) I_{auth} = number of introductions as inferred by the authors.

3) I_{tree} = number of introductions as inferred from a phylogenetic tree with reference sequences included.

4) I_{inh} = number of introduction in the population as calculated using a reference node height distribution.

Table 2

Dating the introductions of local HIV-1 subtype B transmission chains

Molecular clock model	Mean substitution rate (95% HPD) [subst. site ⁻¹ yr ⁻¹]	T _{MRCA} ^{I)} (95% HPD) [yr-month]		
		Stockholm I (N=24)	Stockholm II (N=6)	Sundsvall (N=10)
Strict clock	7.2 (5.3–9.4) × 10 ⁻³	1995–2 (1993–12 – 1995–8)	1997–6 (1996–4 – 1998–6)	1999–9 (1998–9 – 2000–9)
Relaxed clock, lognormal distribution	8.6 (5.3–12.3) × 10 ⁻³	1995–1 (1993–6 – 1995–8)	1997–8 (1996–4 – 1998–9)	1999–9 (1998–8 – 2000–10)
Relaxed clock, exponential distribution	10.7 (6.6–14.8) × 10 ⁻³	1994–11 (1992–7 – 1995–8)	1997–11 (1996–8 – 1999–1)	1999–10 (1998–7 – 2000–11)

^{I)} Mean estimated date of MRCA calculated from last sampling time point for each cluster using BEAST.