Genetic Analysis Reveals Epidemiologic Patterns in the Spread of Human Immunodeficiency Virus

Carla Kuiken,1 Rama Thakallapalli,1 Anne Eskild,2 and Anthony de Ronde3

The extreme variability of human immunodeficiency virus type 1 (HIV-1) makes it possible to conduct transmission studies on the basis of genetic analysis and to trace global and local patterns in the spread of the virus. Two such patterns are discussed in this paper. First, in many European countries (e.g., Scotland and Germany), homosexual men tend to be infected with a subtly different variant of HIV-1 than intravenous drug users. In other European countries (e.g., Norway and Sweden), a distinction is also found between the two risk groups; but based on available data, the distinction is a different one. The second pattern is a worldwide tendency for homosexual men in many different geographic regions around the world to carry HIV-1 subtype B, the variant that is most prevalent in the Americas, Europe, and Australia. In contrast, people infected via other routes (mostly heterosexual contact) in those same countries carry a mixture of other subtypes. Biologic differences between the viruses infecting different risk groups have not been found; the most likely explanation for the findings is different epidemiologic patterns. Although data are still scarce, the authors attempt to use these patterns in the reconstruction of the worldwide spread of the HIV epidemic. Am J Epidemiol 2000;152:814–22.

disease transmission; genetic heterogeneity; genetics; HIV; viruses

The genome of the human immunodeficiency virus (HIV) is extremely variable, both compared with human DNA and compared with most other pathogens. There are two main types of HIV, type 1 (HIV-1) and type 2 (HIV-2). Based on similarity between the viral genomes, HIV-1 is divided into three groups: main (M), outlier (O), and the recently discovered group, new (N). Presently, while groups N and O are not subdivided further, 10 subtypes (labeled A–H, J, and K) of the most common HIV-1 variant (group M) have been recognized. HIV-2 is only subdivided into subtypes (designated by letters, currently A–F).

HIV is now commonly thought to be of zoonotic origin (animal-to-human transmission). HIV-1 is genetically more similar to the chimpanzee strain of simian immunodeficiency virus (SIV), SIV-cpz, than it is to HIV-2. Chimpanzee viruses have been found that are genetically intermediate between HIV-1 groups M and O, as well as between groups M and N, making it plausible that they are related to the common ancestor of both HIV-1 groups (1). Chimpanzees are hunted for food in Africa, and the transmission could have occurred during butchering or eating. The HIV-1 groups M, N, and O probably derive from different chimpanzee-to-human transmissions. The origin of the subtypes that exist within HIV-1 group M is still a matter of speculation. On the other hand, with regard to HIV-2 it is assumed that each subtype corresponds to a new zoonosis, probably from a different species of monkey than that associated with HIV-1. HIV-2 resembles SIV-sm, the sooty mangabey variant of SIV. Sooty mangabeys and monkeys of various other species are kept as pets in Africa, which could explain the transmission of the monkey virus to humans. Only two HIV-2 subtypes (A and B) are epidemic in humans; the others appear not to be transmissible from human to human.

The prevalence of the different HIV-1 group M subtypes depends to some extent on geographic location. Subtype B is by far the most common strain in the Americas, most of Europe, and Australia, but it is very rare in sub-Saharan Africa (except South Africa), the region that is widely regarded as the origin of the HIV epidemic. Of 1,565 African sequences with known subtype in the HIV sequence database at the Los Alamos National Laboratory (Los Alamos, New Mexico), only 27 (1.7 percent) are of subtype B, and these include some infections that can be traced back directly to importation from the United States or Western Europe. In the rest of Africa, subtypes A, C, and D are prevalent, while in West and Central Africa almost all group M subtypes can be found, as well as group N and O viruses.
The prevalence of different subtypes changes with time and with risk category. For example, in the early days of the HIV epidemic in Thailand (around 1985–1990), two subtypes (a B variant dubbed B′ and subtype E) cocirculated, with subtype B′ being found mostly among intravenous drug users (IDUs) and subtype E being most common in heterosexually acquired infections. A few subtype E variants were also found in the Central African Republic, which suggests that subtype E was imported into the Far East from Africa. Subtype E has gained ground rapidly and is now responsible for most new infections among Thai IDUs (19–22). The variability of the virus even that HIV-1 isolated from homosexuals is different from the virus found in IDUs (19–22). The variability of the virus even that HIV-1 isolated from homosexuals is different from the virus found in IDUs (19–22). The variability of the virus even that HIV-1 isolated from homosexuals is different from the virus found in IDUs (19–22). The variability of the virus even that HIV-1 isolated from homosexuals is different from the virus found in IDUs (19–22). The variability of the virus even that HIV-1 isolated from homosexuals is different from the virus found in IDUs (19–22).

In a meta-analysis, we combined and reanalyzed the data on founder effects in HIV-1 subtype B, the most prevalent subtype in Western countries, and further investigated their nature. In this paper, we discuss the implications of these findings for our knowledge of the spread of HIV-1 in the Western world.

MATERIALS AND METHODS

Data sets

This study was based on a large number of existing sets of HIV-1 nucleic acid sequences. Most HIV sequence data are publicly available from the HIV sequence database at the Los Alamos National Laboratory (http://hiv-web.lanl.gov/). The data frequently concern different regions of the HIV-1 genome, so not all data sets can be directly compared. Table 1 presents an overview of the data sets we used.

Methods

Phylogenetic tree analysis is a method of clustering sequences based on their genetic distance (this distance is estimated on the basis of the observed number of differences between the aligned sequences). Phylogenetic trees are used to infer the evolutionary history of sequences and (as in this study) as a tool to graphically represent their genetic differences. In this study, all phylogenetic analyses were carried out using the neighbor-joining algorithm (26), based on a genetic distance matrix calculated using the Kimura two-parameter method (27).

<table>
<thead>
<tr>
<th>Authors, ref. no., and year</th>
<th>Country(ies)</th>
<th>Region of HIV-1 genome</th>
<th>No. of persons</th>
<th>Risk group</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kuiken et al. (29), 1993</td>
<td>The Netherlands</td>
<td>V3*</td>
<td>87</td>
<td>Homosexuals and intravenous drug users</td>
</tr>
<tr>
<td>Albert et al. (19), 1994</td>
<td>Sweden</td>
<td>RT*</td>
<td>21</td>
<td>Homosexuals and intravenous drug users</td>
</tr>
<tr>
<td>de Ronde et al. (unpublished)</td>
<td>The Netherlands</td>
<td>RT</td>
<td>48</td>
<td>Homosexuals and intravenous drug users</td>
</tr>
<tr>
<td>Engelstad (22), 1996</td>
<td>Norway</td>
<td>V3</td>
<td>22</td>
<td>Homosexuals and intravenous drug users</td>
</tr>
<tr>
<td>Lukashov et al. (31), 1996</td>
<td>The Netherlands and the United States</td>
<td>V3</td>
<td>72</td>
<td>Homosexuals and intravenous drug users</td>
</tr>
<tr>
<td>Brown et al. (30), 1997</td>
<td>Scotland and Ireland</td>
<td>V3</td>
<td>211</td>
<td>Homosexuals and intravenous drug users</td>
</tr>
<tr>
<td>Kuiken et al. (20), 1996</td>
<td>The Netherlands, Germany, and Scotland</td>
<td>V3, vpr, and vpu</td>
<td>76</td>
<td>Homosexuals and intravenous drug users</td>
</tr>
<tr>
<td>Estable et al. (41), 1998</td>
<td>Canada</td>
<td>V3</td>
<td>84</td>
<td>Homosexuals and intravenous drug users</td>
</tr>
<tr>
<td>Quinones-Mateu et al. (35), 1996</td>
<td>Spain</td>
<td>RT</td>
<td>26</td>
<td>Homosexuals and intravenous drug users</td>
</tr>
<tr>
<td>Quinones-Mateu et al. (61), 1995</td>
<td>Venezuela</td>
<td>V3</td>
<td>8</td>
<td>Homosexuals and heterosexuals</td>
</tr>
<tr>
<td>Scarlatti et al. (37), 1993</td>
<td>Italy</td>
<td>V3</td>
<td>9</td>
<td>Intravenous drug users</td>
</tr>
<tr>
<td>Zheng et al. (62), 1996</td>
<td>Australia</td>
<td>RT</td>
<td>24</td>
<td>Homosexuals</td>
</tr>
<tr>
<td>Nerurkar et al. (43), 1995</td>
<td>Hawaii (United States)</td>
<td>V3</td>
<td>26</td>
<td>Homosexuals and intravenous drug users</td>
</tr>
<tr>
<td>Adwan et al. (38), 1999</td>
<td>Greece</td>
<td>V3</td>
<td>18</td>
<td>Homosexuals and heterosexuals, and intravenous drug users</td>
</tr>
<tr>
<td>Casado et al. (36), 1999</td>
<td>The Netherlands, Germany, Italy, France, and Spain</td>
<td>V3</td>
<td>37, 11, 17, 46, and 69</td>
<td>Homosexuals and intravenous drug users</td>
</tr>
<tr>
<td>Kim et al. (59), 1999</td>
<td>Korea</td>
<td>V3</td>
<td>58</td>
<td>Various risk groups</td>
</tr>
</tbody>
</table>

* V3, third variable region of the envelope gene; RT, reverse transcriptase.
Principal coordinate analysis, an adaptation of principal component analysis for sequence data, was performed using the PCOORD program (28). The method summarizes the total variation in the matrix of pairwise distances between the sequences in a limited number (maximum 10) of dimensions or “coordinates.” A coordinate is a pattern in the data, usually caused by correlated changes in multiple positions. Principal coordinate analysis can be used to find groups of sequences that show correlated mutations.

The statistical significance of the mutational patterns was determined as described elsewhere (20). Briefly, the actual distribution of mutations among groups was compared with the distribution found in 100 simulated groupings over all positions in the sequence, and a distribution that was more extreme than the most extreme one found in those random groupings was considered significant.

RESULTS

The United States and Europe

A distinction between viruses from different risk groups in the Netherlands has been described previously (29) (an overview of the mutations associated with each country and risk group can be found in table 2). The distinction was very consistent: 146 out of 150 persons could be assigned to the correct risk group on the basis of the most consistent mutation. Four IDUs had a virus with non-IDU characteristics. Of these individuals, one was found upon closer investigation to have dual risk and one had died of unrelated causes, which made it impossible to gather further information. For another individual, the discrepancy was found to have resulted from a sample mix-up. For the fourth person, no explanation for the aberrant pattern could be found. All six persons with hemophilia who were studied were infected with the same variant as homosexuals. Since blood donation in the Netherlands is an unpaid community service, IDUs there rarely donate blood. A later study comparing different regions of the viral genome found a similar distinction in Dutch, German, and Scottish samples (20). The distinction in Scotland was later confirmed and extended in a larger study that also comprised British and Irish samples (30).

It has been suggested that the HIV-1 variant found among IDUs in these countries originated in the United States, where approximately half of the infected IDUs from a Baltimore, Maryland cohort were found to be infected with a similar variant (31). Studies on the geographic mobility of IDUs in Edinburgh, Scotland showed frequent needle-sharing in other Scottish cities, as well as some in the Netherlands, Southern Europe, and the United States (32, 33).

For this study, we compared virus variants isolated from both risk groups in Sweden (reverse transcriptase sequences (19)) and Norway (sequences from the third variable (V3) region of the envelope gene (22)). The Dutch and Swedish sequences depicted in figure 1 show two types of significant distinctions. The first (indicated by solid arrows) are positions that distinguish IDUs (both Swedish and Dutch) from homosexuals from those two countries. The second type, indicated by open arrows, distinguishes between one risk group from one country (e.g., Swedish IDUs) and all other groups. Based on this set of sequences, it can be seen that in Sweden, as in the Netherlands, Scotland, and Germany, different risk groups carry distinct HIV-1 variants. Some of the

<table>
<thead>
<tr>
<th>Comparison</th>
<th>Region of HIV-1 genome</th>
<th>Position numbers in HxB2*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Homosexuals and intravenous drug users in the Netherlands, Scotland, and Germany</td>
<td>V3†</td>
<td>7046 (T/C), 7072 (C/T), 7166 (G/C), 7242 (A/G)</td>
</tr>
<tr>
<td>Homosexuals and intravenous drug users in the Netherlands, Scotland, and Germany</td>
<td>vpr</td>
<td>5591 (G/A), 5613 (A/G), 5789 (G/A), 5816 (A/G)</td>
</tr>
<tr>
<td>Homosexuals and intravenous drug users in the Netherlands, Scotland, and Germany</td>
<td>vpu</td>
<td>6235 (A/C), 6271† (G/A), 6299† (G/A), 6309 (G/A), 6323 (A/T)</td>
</tr>
<tr>
<td>Homosexuals and intravenous drug users in the Netherlands</td>
<td>RT†</td>
<td>2660 (T/C), 2717 (C/T), 2797 (G/A), 3159 (T/C)</td>
</tr>
<tr>
<td>Homosexuals and intravenous drug users in Sweden</td>
<td>RT</td>
<td>2695 (A/G), 2717 (C/T), 2797 (G/A), 3159 (T/C)</td>
</tr>
<tr>
<td>Homosexuals and intravenous drug users in Norway</td>
<td>V3</td>
<td>7048 (A/C), 7091 (A/C), 7114 (C/T), 7261 (T/C), 7264 (C/T)</td>
</tr>
<tr>
<td>Homosexuals and intravenous drug users in Canada</td>
<td>V3</td>
<td>7086 (C/T), 7140 (A/G), 7160 (A/G), 7179 (A/G), 7208 (A/G), 7221 (A/C), 7233 (A/C), 7272 (G/A)</td>
</tr>
</tbody>
</table>

* Underlined nucleotides form the majority among the subtype B sequences deposited in the Los Alamos National Laboratory HIV sequence database. The first nucleotide given is associated with the homosexual group.
† V3, third variable region of the envelope gene; RT, reverse transcriptase.
‡ Not significant in the Scottish and German samples (n = 14).

Am J Epidemiol Vol. 152, No. 9, 2000
Genetic Analysis of Patterns in the Spread of HIV

FIGURE 1. Schematic representation of the alignment of Swedish (19) and Dutch (de Ronde et al., unpublished data) reverse transcriptase (RT) sequences of human immunodeficiency virus type 1 from homosexuals (HS) and intravenous drug users (IVDU). Positions where the sequences differ from the consensus (a sequence that contains the most common carrier at each position) are shown as square dots. The consensus characters are indicated above the arrows. (T, thymine; A, adenine; C, cytosine; G, guanine.) It can easily be seen that at several positions there are columns of mutations coinciding with risk group in both countries (solid arrows) or in one country (open arrows). The percentage of sequences with each mutation is shown in italic type.

The distinguishing positions are the same for both countries, while others are unique to either the Dutch population or the Swedish population. This pattern (a number of shared mutations and a smaller number of mutations that are unique to either group) suggests separate transmissions from one outside population, in which the mutations that are shared by Swedish and Dutch IDUs are common while those that are unique to one country are relatively rare.

Unfortunately, we did not succeed in finding either a well-documented set of reverse transcriptase sequences from US IDUs or a sizeable set of Swedish V3 sequences from well-defined risk groups, so no comparisons could be made with viruses prevalent among US IDUs.

The Norwegian case can be illustrated by means of a principal coordinate plot. Principal coordinate analysis is very similar to principal component analysis. Cases (sequences) are grouped together on the basis of their scores on a number of dimensions called principal coordinates. A principal coordinate is any important source of variation in the data, and usually consists of a group of positions that behave similarly (e.g., if position 230 contains an A, position 385 often contains a C). Figure 2 compares the Norwegian sequences...
with the Dutch and Baltimore sequences. All Norwegian IDUs except one lie outside of the Dutch/US IDU cluster. Instead, they are positioned on the fringe of the main (homosexual) group. The sequences from Norwegian and Dutch homosexuals appear to be indistinguishable.

The one Norwegian IDU who carried the “Dutch IDU variant” attributed his infection to contact (sexual contact and needle-sharing) with a woman who had traveled to the West Indies. He was infected in 1984 and thus was one of the first infected IDUs inside Norway; the onset of the HIV epidemic among Norwegian IDUs has been dated at 1983, based on analysis of stored serum samples (34).

We can conclude that differences between homosexual and IDU virus populations appear to be ubiquitous in Northern Europe. The variant found among homosexuals in all cases examined is the most common variant in the Los Alamos HIV sequence database. This could be a sampling effect: Since homosexuals in the Western world tend to be more motivated and more reliable study participants, many HIV strains are derived from this population.

In Southern European countries, the situation is less clear. In a set of reverse transcriptase and V3 sequences from Spain (35), no trace of a risk group-related distinction in viral variants could be found. The same thing was found in a second study that also included sequences from France, where the risk groups also appeared to carry indistinguishable variants (36). In Italy, on the other hand, a recent study of 17 V3 sequences from IDUs and homosexual men indicated that the distinction there is preserved (36). In a mother/infant transmission study among Italian IDUs (37), four patients carried the IDU mutation, while five did not. A recent data set of 12 V3 sequences from Greece (38) did show a clear distinction between IDUs and others that resembled the one found in the Netherlands.

The United States and Canada

Large sets of sequences from the United States with mixed and well documented risk groups are scarce. We were able to find a small set of V3 sequences from nine New York City individuals; the V3 mutation at position 7166 was found in all three IDU viruses but in none of six hemophilia patients (39, 40). As was noted above, about half of the infected IDUs from the Baltimore cohort were infected with the same IDU-associated variant. The fact that (based on the limited data available) the distinction appears to be present in two cities that are approximately 400 miles (640 km) apart makes it tempting to speculate that this epidemiologic pattern may be found throughout the urban northeastern United States. More data are needed to assess the validity of this hypothesis.

There is a well documented data set from Vancouver, Canada (41) that shows very clear differences between sequences from IDUs and homosexuals. Once again the sequences are of the V3 region, which allows comparison with the Dutch and Norwegian risk group distinction but not with the Swedish one. The distinctive positions in Vancouver are not the same as the ones found in Norway or the Netherlands. In a smaller set of V3 sequences from rural Georgia (42), there was no distinction between homosexuals and heterosexuals; the set contained no IDUs. In a set of V3 sequences from Hawaiian IDUs and heterosexually infected persons (43), the mutation most consistently associated with intravenous drug use in Amsterdam (HxB2 position 7196) was found in two of the 10 IDUs and in none of the 16 heterosexuals. The mutations seen in virus from drug users in Norway were also rare or absent in the Hawaiian sequences, as were the drug user mutations found in Vancouver. It seems reasonable to conclude that the virus in this population of IDUs is unrelated to the one found in Amsterdam, Baltimore, and New York, and that different populations of IDUs around the United States carry different virus variants.

Based on epidemiologic studies, the most likely source of infection for the heterosexual population in the Western world (at least for those who were infected through local sexual contact) is intravenous drug use. In a modeling study based on US data, it was estimated that approximately 1 percent of infections in heterosexual women result from sexual intercourse with bisexual men (44). Studies based on self-report data from HIV-1-infected bisexual men in San Francisco, California show very infrequent risk contacts with women (45). Similarly, in the Dutch cohort study of homosexual men, very few participants indicated ever having had sex with a woman, and most of those men always used condoms (46). A second study indicated that homosexual prostitution by male IDUs also is not a significant factor for introduction of HIV into the heterosexual population (47). In this light, it is perhaps surprising that more than half of a group of heterosexually infected women in the Netherlands carried the “homosexual” variant (48). Based on comparisons of samples taken over a period of 10 years, the prevalence of the IDU variant in Dutch drug users does not seem to be declining (49). However, as noted previously, the homosexual variant forms the vast majority of subtype B sequences collected. It may be that this variant forms the majority among present-day HIV-1-infected heterosexuals.
in the United States, and some of the Dutch infections are imported from abroad. Alternatively, the frequency of unprotected bisexual contact (the “bisexual bridge”) may be underestimated by epidemiologic investigations, which are often based on interviews with male participants in HIV studies who are committed to the homosexual lifestyle.

The rest of the world

HIV found in samples obtained from homosexuals often belongs to subtype B, even when the locally prevalent subtype is a different one. The distinction has been reported in three studies so far: one from the former Soviet Union (50), one from South Africa (51), and one from Thailand (52). A fourth report from Singapore is currently in preparation (M. Kalish, Centers for Disease Control and Prevention, personal communication, 1999). These results suggest that the HIV epidemic among homosexual men is evolving separately from that among IDUs and heterosexuals, not only in the West but also in other parts of the world. Even before the genetic difference between the HIV variants was known, it was suggested that the HIV epidemic among homosexuals in South Africa was probably imported from Europe or the United States (53).

Data from East Asia (Taiwan, Korea, and Japan) suggest that a large fraction of infections there are caused by subtype B (54, 55). In a recent Japanese study, infections among homosexuals and persons with hemophilia were exclusively of subtype B, while heterosexually infected persons carried other subtypes (54). An earlier study of five hemophilia patients showed that, similar to the Dutch finding, they were infected with the strain circulating among homosexuals, in this case US-like subtype B (56). Recently, several Korean data sets (nef and V3 sequences) were published, consisting of samples from patients infected through heterosexual and homosexual contact and infected blood products (57–59). In one study, 41 of 46 isolates were of subtype B (57); infections with other subtypes were invariably associated with heterosexual transmission. Sequences in all three Korean studies showed the presence of an extreme founder effect in Korea (57, 58); the Korean variant of subtype B is quite distinct from the worldwide consensus. This can be seen in the phylogenetic tree presented in figure 3. Phylogenetic trees graphically represent the genetic similarity between sequences, so that more similar sequences are grouped together. The tree in figure 3 is based on V3 sequences from Korea and Hawaii (Hawaii is also a small and geographically fairly isolated region), as well as representatives of various other HIV-1 subtypes. Aside from five subtype A sequences and seven representatives of “US-type” subtype B sequences, most (24/36) of the Korean viruses cluster tightly together. They are clearly distinct from the main subtype B, as well as from the Thai B′ variant. In contrast, all Hawaiian sequences are mixed with the other subtype B sequences. None of the Korean studies found a distinction between risk groups. Kim et al. (59) reported that many homosexual men in Korea also have sex with women, which may explain the mixing of variants between the homosexual and heterosexual populations.

![Phylogenetic tree of human immunodeficiency virus type 1 V3 sequences](image)

**Figure 3.** Phylogenetic tree of human immunodeficiency virus type 1 V3 sequences (third variable region of the envelope gene) from Hawaii (43) and Korea (59). In this radial tree, only the lengths of the branches connecting the endpoints have meaning; the actual positions of the endpoints are meaningless. Branches labeled with letters represent the major subtypes (A–G) of human immunodeficiency virus type 1 and are included for reference. The scale bar represents a 5% genetic difference between two sequences connected by a branch of that length.

In several data sets from South America (60, 61), no differences were seen between sequences derived from the homosexual population and those derived from the IDU population, but the data were frequently sparse and the risk group of the patients was often unknown. In a Brazilian study, no subtype distinction between the risk groups was found (60). However, more than 80 percent of the subjects in this study were infected with subtype B; since the typical pattern of subtype distinction involves subtype B in homosexuals versus the endemic subtype in other risk groups, this pattern would be invisible in a country where subtype B is the endemic subtype. Unfortunately, very few sequences were generated for this study. Most of the analyses were conducted using a technique called “heteroduplex mobility assay,” which only allows comparison with a supplied set of viral strains and precludes further analysis of these data to search for risk group-associated patterns. A second South American data set consisting of Venezuelan V3 sequences from homosexually and heterosexually infected patients showed no difference between the two groups (61).

A set of reverse transcriptase sequences from 24 Australian homosexual men (62) showed that none of them had any of the mutations that characterize either Dutch or Swedish IDU sequences. For all of these positions, the homosexual variant again formed the majority in reverse transcriptase sequences deposited in the Los Alamos database.

**DISCUSSION**

The data presented above allow us to sketch a tentative scenario for the introduction of HIV-1 into the Western world. Based on the available data, the most plausible scenario seems to be that the subtype B virus was carried out
of Africa and introduced into the Western homosexual community by one person. There is some other evidence for this in the form of the notorious “patient 0,” a gay airline steward who was infected in the late 1970s and allegedly infected a number of his sex partners (63). A sequence of this patient’s virus is available; it is of subtype B and has the homosexual characteristics. However, patient 0 probably was not solely responsible for the initial spread of the virus in the United States; anecdotal evidence shows that he was part of a cluster of homosexual men who traveled frequently, were extremely sexually active, and died of acquired immunodeficiency syndrome at a very early stage in the epidemic (around 1980–1982) (64).

An alternative possibility consistent with these data is that the virus arrived in the United States via Haiti. The first Haitian acquired immunodeficiency syndrome patient was identified in 1978, the same year the disease was first seen in the United States, but the prevalence of HIV in Haiti turned out to be much higher (65). Two scenarios are plausible: Either the virus was introduced into Haiti from the United States at a very early stage in the epidemic, possibly through gay sex tourism (Haiti was a popular holiday destination for gay men (66)), or the virus was introduced into Haiti via African contacts, possibly Haitian migrant workers in Congo/Zaire (67), and spread to the US homosexual population via sex tourism. Since the vast majority (85 percent) of Haitian patients were male (68), it is now commonly assumed that homosexual contacts formed the “HIV bridge” between the two countries. The Haitian virus is also of subtype B, and all 82 Haitian sequences presently available have the homosexual signature at the main position in V3.

From the United States, the virus presumably spread from one Western country to the next, most likely via homosexual contact. There are some isolated early cases of suspected HIV in Europe: a Norwegian sailor and his family, infected with the African group O virus, and two possible but not experimentally confirmed cases in homosexual men in Germany (69) and Austria (70) (also see Hooper (64)). However, the epidemic did not gather momentum in Europe until several years after it appeared in the United States. In 1982, the prevalence of HIV among homosexuals was 42.6 percent in San Francisco and 7.5 percent in Amsterdam (71).

The first acquired immunodeficiency syndrome patients in the United States and Western Europe (and in most other countries where subtype B predominates, such as Australia (72), Taiwan (73, 74), and Korea (57)) were homosexual men. This strongly suggests that the virus was first introduced into this risk group and spread to other risk groups from there. Because subtype B is so rarely found in Africa, the odds of a second independent introduction from Africa with the same rare subtype are very low. In addition, estimates indicate that the epidemic among IDUs in a given country usually lags several years behind that among homosexuals. Aside from epidemiologic evidence (75, 76), this hypothesis is supported by genetic data. In figure 4, the variability in sequences from the IDU and homosexual populations in various European countries is compared; in all cases, the IDU virus is more homogeneous than the virus isolated from homosexual men. Since the genetic variability of the virus generally increases with time in both populations (29, 49), this again suggests that the IDU epidemic lags behind the homosexual epidemic.

As is shown above, several variants of the virus circulate among IDUs across Europe, indicating multiple introductions from the subtype B pool into this risk group. The presence of supranational IDU-related mutations in European countries, which in each case form an exception to the US-based “normal” subtype B, suggests that the virus in these countries was not derived from the local homosexual epidemic but rather spread from a separate source or sources. On the basis of numbers alone, it would be expected that the epidemic among IDUs would also derive from the United States; but presently the only real evidence for such an origin is the finding, in Baltimore and in a small set of sequences from New York, of a variant that contains the same mutations with respect to the universal subtype B consensus as the Northern European IDU sequences. After it arrived in Europe, the virus presumably spread from one European country to another. International travel and needle-sharing among European IDUs is well documented (30, 39, 77).

Sequencing of more virus variants from a large number of US IDUs from different geographic locations would seem to be the best way to determine the plausibility of this scenario. Support for this hypothesis would be provided if a group of IDUs who carried the mutations found in virus from Norwegian and/or Swedish IDUs could be found.

ACKNOWLEDGMENTS

This study was assigned LAUR number 99-2437. Part of this work was funded through the Laboratory Directed Research and Development Program (Delphi Project) of the Los Alamos National Laboratory.

The authors thank Drs. Bette Korber and Bill Bruno of the Los Alamos National Laboratory for helpful advice and discussion.
REFERENCES


Retroviruses 1998;14:917–19.


