The Use of Molecular Evidence in the Investigation of Human Immunodeficiency Virus (HIV) Transmissions

Nakita Ashanta Haynes, Trent University

The high genetic variability of the Human Immunodeficiency Virus has been the driving force for its widespread use in the investigation of forensic cases involving HIV transmissions. As a consequence of this genetic variability, HIV variants arise that are unique to individuals hence viruses isolated from those infected by the same source would possess a considerable measure of genetic homogeneity. There are currently three HIV genes that are used as genetic markers in such forensic investigations. The sequences of these genes are used to assess the genetic relatedness between viruses isolated from suspected transmission pairs. Thus, molecular evidence can be used to establish HIV transmission links between a suspected donor and recipient. This paper aims to present a review of the HIV genetic markers used in HIV transmission investigations. Emphasis will be placed on the characteristics of the markers, the techniques used in their analysis and the limitations associated with their use. Future directions in forensic HIV transmission investigations will also be discussed.

1. The biology of HIV

The Human Immunodeficiency Virus (HIV) is an enveloped particle with a genome consisting of two copies of single-stranded RNA [1]. These RNA molecules are protected by two protein layers (capsid and matrix), which are surrounded by a lipid envelope containing the proteins gp120 and gp41 (Figure 1). The envelope proteins allow the virus to penetrate the host cell [2].
The Use of Molecular Evidence in the Investigation of Human Immunodeficiency Virus (HIV) Transmissions

FIGURE 1: The structure of an HIV-1 virion. The virion consists of a single-stranded RNA genome enclosed by a layer of capsid proteins. A lipid envelope surrounds the virion and has a number of associated envelope and matrix proteins [3].

HIV is a member of the family of viruses called Retroviridae (genus: Lentivirus), which are characterized by the ability to form a DNA copy of their RNA genome upon entry into the host’s immune cells [1]. This genome copy is called proviral DNA and is manufactured by a reverse transcriptase enzyme encoded on the viral RNA. The virus then integrates the proviral DNA into the genome of the host cell, which initiates the replication cycle thereby allowing for the formation of new viral particles (Figure 2) [1].

FIGURE 2: The HIV life cycle. The virus enters the host cell and manufactures a proviral DNA copy of its RNA genome, which is subsequently integrated into the genome of the host cell. The virus is then able to replicate [4].
Two species of HIV exist: HIV-1 and HIV-2. Both species can be transmitted sexually and through contact with infected blood or bodily fluids. However, HIV-1 and HIV-2 are considerably dissimilar as only 43% of their DNA sequences are identical [5]. Additionally, HIV-2 is not widespread as it is found only in Africa, it is less easily transmitted and shows reduced virulence compared to HIV-1 [5]. Conversely, Grez et al. (1994) found that 21% of HIV infections in Bombay, India were of the HIV-2 species, which suggests that HIV-2 can in fact be easily transmitted within a distinct geographical location [6]. HIV transmissions are a global concern as it is estimated that approximately 36 million individuals are infected worldwide [7].

2. The forensic use of HIV molecular evidence

HIV transmissions can potentially develop into legal matters when they occur as a result of criminal activities (such as rape and intentional transmissions from an infected individual to a sexual partner) as well as civil situations involving medical malpractice (for example, a hospital patient being administered infected blood or blood products) [8]. In such instances a scientific expert would be required to uncover evidence that would either support or rebut a transmission link between the suspected donor and recipient [8]. It can be seen why such evidence is necessary when one considers the fact that a high degree of inaccuracy has been found when individuals identify their presumed source of infection. For example, Trask et al. (2002) reported that out of a sample of 149 couples who reported transmitting the virus to each other, 13% of transmissions were refuted by molecular evidence [9].

Numerous cases are reported in scientific literature involving the investigation of HIV transmissions using molecular evidence. In these cases an epidemiological link was ascertained by verifying the genetic relatedness between viruses isolated from the suspected transmission pair [9]. Although approximately 75% to 80% of HIV transmissions are by way of sexual contact, the cases most commonly discussed in literature are those involving nosocomial transmissions of the virus from an infected medical practitioner to patients.
under his or her care [10]. A well-known case is that involving a HIV-infected dentist from Florida who was found to be the source of infection of a number of his patients after it was discovered that the viruses from the patients were similar to those from the dentist [11]. Such cases may be even more significant when it is established that a medical practitioner failed to comply with the infection control protocols stipulated by the relevant authorities [12].

The use of HIV molecular evidence had found widespread acceptance in criminal courts of law after the legal precedent was set by the Louisiana case involving a physician who deliberately infected his former girlfriend with the blood of one of his HIV-infected patients [13]. In this case the use of HIV molecular evidence was found to meet the legal criteria for admissibility into court in that the techniques are subjected to experimental testing, peer review and publication, the error rates can be established and the methods are generally accepted within the scientific community [13]. Consequently, the study of HIV transmissions is not only important from an epidemiological perspective where the focus is to track the disease in an attempt to identify emerging strains; but it can assist criminal prosecutions, civil lawsuits and even divorce cases by identifying potential sources of infection.

3. HIV genetic markers

The HIV genome consists of nine genes, which are encoded on a single-stranded RNA molecule [2]. Upon entry into the host cell, the virus synthesize a 9.5 kilobase proviral DNA copy of its RNA genome, which is then integrated into the genome of the host (as is seen in Figure 2). The nine genes contain the information necessary for the regulation of the virus’ life cycle as well as its ability to cause disease (Figure 3) [2].
FIGURE 3: The 9.5 kilobase proviral HIV genome. The genome consists of nine genes that encode information necessary for the regulation of the virus’ life cycle as well as its ability to cause disease. The three genetic markers used in HIV transmission investigations are highlighted in yellow [14].

Three of the nine HIV genes are commonly used to assess the degree of genetic relatedness between viruses isolated from a suspected donor and recipient. These genes are env, gag and pol. Env is the approximately 2.5 kilobases envelope gene, which encodes the gp120 and gp41 proteins of the viral envelope [2,15]. Gag or the group specific antigen gene is roughly 1.4 kilobases in size and encodes the protective matrix and capsid proteins [2,15]. Pol is the approximately 2.8 kilobases polymerase gene that encodes enzymes such as reverse transcriptase, which are essential for viral replication [2,15]. Within env, gag and pol there are specific regions that are also used in forensic investigations as they show signs of substantial genetic variability. These include the gp120 and V3 regions of env, the p17 region of gag and the RT region of pol. Refer to Table 1 for the size of each genetic marker.

TABLE 1: The size of the genetic markers used in HIV transmission investigations

<table>
<thead>
<tr>
<th>GENETIC MARKER</th>
<th>SIZE (nucleotides)</th>
<th>REFERENCE</th>
</tr>
</thead>
<tbody>
<tr>
<td>env</td>
<td>2,529</td>
<td>Li et al. (1992) [15]</td>
</tr>
<tr>
<td>gp 120</td>
<td>1,533</td>
<td>Kviken et al. (2008) [16]</td>
</tr>
<tr>
<td>V3</td>
<td>105</td>
<td>Milich et al. (1992) [17]</td>
</tr>
<tr>
<td>gag</td>
<td>1,398</td>
<td>Li et al. (1992) [15]</td>
</tr>
<tr>
<td>p17</td>
<td>396</td>
<td>Kviken et al. (2008) [16]</td>
</tr>
<tr>
<td>pol</td>
<td>2,841</td>
<td>Li et al. (1992) [16]</td>
</tr>
<tr>
<td>RT</td>
<td>1,320</td>
<td>Kviken et al. (2008) [13]</td>
</tr>
</tbody>
</table>
Env, gag and pol have found widespread forensic use for three main reasons. Firstly, they are present in all HIV strains, which make their use universal in the analysis of all suspected transmission links [13]. Secondly, these genes contain conserved regions, which have proven to be beneficial for the use of techniques such as PCR where universal primers can be designed to anneal to specific regions within the HIV genome [18]. Most importantly, env, gag and pol contain variable domains that confer a high degree of genetic uniqueness between viruses from different individuals. In particular, env is the most variable while pol is the least [19,20]. By this means, in instances where a transmission link exists, HIV variants from transmission pairs are expected to exhibit great similarity when compared to isolates from unrelated individuals [18].

The genetic variability of env, gag and pol is the result of a number of factors. Firstly, within the host the virus is under a considerable amount of environmental pressure from both the host’s immune system as well as any antiretroviral drugs within the blood stream. Therefore, to avoid destruction the virus must evolve rapidly in response to such pressures. For example, env and gag sequences evolve at an annual rate of 1% and 0.5%, respectively [21]. Accordingly, HIV has a high mutation rate of approximately one error for every one thousand to ten thousand bases incorporated during replication [22,23]. Moreover, the high viral replication rate (approximately ten billion virions are produced per day) along with genetic recombination during replication favours the insertion, deletion and reshuffling of genetic material [24]. Specifically, indels are often less favoured than substitutions; adenine to guanine transitions are most common [25].

4. The analysis of HIV genetic markers

Prior to the analysis of HIV genetic markers, blood samples are obtained from the suspected donor and recipient plus controls of HIV infected individuals from the same geographic location and associated risk groups. The use of local controls is important because it allows for the identification of
an alternative transmission route [13]. Additionally, as molecular studies have
determined that distinctive HIV variants are found in specific geographic
locations and risk groups, the examination of control sequences can determine
if the suspected transmission pair possesses variants that are prevalent within
their respective locales [13,26].

A number of polymerase chain reaction (PCR)-based techniques are
used to examine the genome of viruses isolated from different individuals.
These techniques include: restriction fragment length polymorphism (RFLP)
analysis and oligonucleotide probe hybridization. However, DNA sequencing
is the most accurate and is therefore the most widespread [27].

After the samples are collected, DNA is extracted and a fragment of the
gene of interest is amplified by PCR. The amplicons from each sample are
then sequenced and the sequences are compared. If the viruses are from the
same source it is expected that the sequences would be similar. For example,
genetically related $env$ sequences exhibit between 95% to 100% similarity
[28,29]. Viral sequences can also link a suspect to a victim when unique
mutations are observed. Such is the case where it was found that both viral
isolates contained a three nucleotide out-of-frame deletion in the $gag$ gene
[30]. In another case, the presence of a specific mutation in the victim’s viral
$RT$ region allowed investigators to determine that the source possessed a
variant that was resistant to the antiretroviral drug azidothymidine (AZT)
and therefore must have been undergoing drug therapy [13].

However, as a result of the rapid evolutionary rate of HIV the assessment
of the genetic similarity between viral isolates does not merely involve a simple
comparison of the relevant DNA sequences. The level of variation between
HIV gene sequences is dependent on the amount of time that has passed since
the transmission occurred thus phylogenetic analyses must be employed to
establish evolutionary relationships [19,31]. In general, one phylogenetic tree
is constructed for each genetic marker analyzed and the trees are compared.
This is necessary because each gene is under different environmental pressures
and therefore undergoes divergent evolutionary rates. Thus, comparative
phylogenetic analyses strengthen the validity of findings as was found by
The Use of Molecular Evidence in the Investigation of Human Immunodeficiency Virus (HIV) Transmissions

Leither et al. (1996), who reported achieving more accurate results when both the $p17$ region of $gag$ and the $V3$ region of $env$ were used in contrast to the use of one marker alone [32].

The probability of transmission from the suspected source to the victim is determined by statistical likelihood analysis where the genetic similarity of viral sequences from the individuals in the case is compared against the control sequences [18]. If the viral sequences are from the same source then they would cluster together on the phylogenetic tree and form a lineage that is separated from the control sequences as is seen in Figure 4.

![Phylogenetic Tree](image)

**FIGURE 4:** A phylogenetic tree showing the evolutionary relationship between viral DNA sequences from the RT region of $pol$ that were isolated from a suspected transmission pair. The victim’s gene sequences are embedded within those of the suspected source and both form a lineage that is separated from the control sequences [13]. (Adapted from Metzker et al., Molecular evidence of HIV-1 transmission in a criminal case, PNAS, 99(22) (2002) 14292-14297).

5. Issues surrounding the use of HIV molecular evidence

A number of central issues surround the acquisition of HIV molecular evidence for the use in transmission investigations. The first concern is with the time that the samples are taken. Immediately after the transmission event, the
recipient’s viral population is identical to that of the source but as the infection advances, the virus evolves in response to the different host environment. Consequently, variants arise that are genetically distinct from the source [19]. For instance, V3 sequences are only conserved during the first 24 weeks after the infection occurred [33]. Thus, samples must be obtained near the time of transmission to increase the likelihood of identifying epidemiologically linked cases [19]. Lymphocytes are best for this purpose as they are known to retain ancestral sequences [34]. This situation is even more complex when one considers the fact that several variants may be present within one individual. Therefore, samples must represent the diversity of the gene within the host [13].

There is also a concern with the validity of the control group used in analyses. For instance, if the control group contains HIV subtypes that are not related to the forensic case then the viral isolates from the suspected transmission pair would appear to be more closely related to each other than to the controls, even in situations where a transmission link is nonexistent [35]. Moreover, the interpretation of results may be confounded in instances where there is a rapid outbreak of a particular variant. A transmission link could potentially be overlooked since the sequences from the transmission pair would closely resemble the control sequences. Thus, the control group must be carefully assembled and samples should also be obtained close to the time of the transmission event [19].

Another issue surrounds the genetic marker chosen for analysis given that some markers have been found to be more informative than others. For example, the p17 region of gag is very small and may not be variable enough to assess genetic relatedness. Conversely, markers such as the V3 region of env may be too variable [18]. Stürmer et al. (2004) recommends analysing two or more regions from different genes so as to increase the reliability of results [31].

There are also limitations associated with the forensic use of HIV molecular evidence. Firstly, if a transmission link is found between the suspected donor and recipient the direction of transmission cannot be determined. Also, the
possibility cannot be eliminated that another individual infected both persons involved in the case [35]. Thus, HIV molecular evidence cannot be used as the only proof that a transmission link exists and can only exclude suspects or corroborate other evidence in a case [19].

6. Future directions in HIV transmission investigations

Although DNA sequencing is the most accurate and informative method used to examine the genetic relatedness between HIV variants, there has been a need to develop more rapid and cost-effective techniques. One such technique is heteroduplex mobility assay (HMA) [36]. It is based on the principle that once related viral sequences are allowed to anneal, heteroduplexes are formed which have reduced mobility during polyacrylamide gel electrophoresis. The greater the similarity between viral sequences the larger the extent at which the heteroduplex is retarded on its migration through the gel [36]. Therefore HMA can be potentially used as a screening tool to determine if the HIV variants under question are similar, before more extensive investigations involving DNA sequencing and phylogenetic analyses are conducted.

It is hoped that in the future the degree of genetic diversity between HIV variants can be assessed in efforts to determine the time of infection. In addition, HIV transmission investigations should involve the analysis of both viral RNA as well as proviral DNA sequences in order to attain a more accurate estimation of viral diversity [37].

References


[16] C. Kuiken, T. Leitner, B. Foley, B. Hahn, P. Marx, F. McCutchan, S. Wolinsky, B. Korber,


[29] C. Kuiken, B. Korber, Epidemiological significance of intra- and inter-person variation


