Recombinant HIV Sequences: Their Role in the Global Epidemic

Martine Peeters

Laboratoire Retrovirus, IRD, 911 Avenue Agropolis, BP5045 34032 Montpellier, France
E-mail: martine.peeters@mpl.ird.fr

INTRODUCTION

One of the major characteristics of the human immunodeficiency viruses (HIVs) is their extremely high genetic variability. This extensive heterogeneity is the result of the high error rate of reverse transcriptase (76), and the fast turnover of virions in HIV-infected individuals (32, 107). In addition, the reverse transcriptase enzyme is known to be highly recombinogenic (35), so that radically different genomic combinations may be generated in individuals infected by genetically diverse viruses. Recombination requires the simultaneous infection of a cell with two different proviruses, allowing the encapsidation of one RNA transcript from each provirus into a heterozygous virion. After the subsequent infection of a new cell, the reverse transcriptase, by jumping back and forth between the two RNA templates, will generate a newly synthesized retroviral DNA sequence that is recombinant between the two parental genomes (28, 35, 94). That mosaic viruses are indeed recombinants is supported by the fact that discrete breakpoints can be identified between the genomic regions with different phylogenetic associations (13, 23). It is now well established that recombination is a relatively common occurrence among different strains of HIV (reviewed in 77). Recombination is most obvious among members of different subtypes, and is also likely to occur among members of the same subtype, although current methods fail to reliably identify such intra-subtype recombination.

CLASSIFICATION OF HIV-1 STRAINS

Phylogenetic analyses of numerous strains of HIV-1, isolated from diverse geographic origins, have revealed that they can be subdivided into groups, subtypes, sub-subtypes and CRFs (81, 82, 83, 103), see Figure 1.

Figure 1: Evolutionary relationships among non-recombinant HIV-1 strains from the HIV-1/SIVcpz lineage, based on neighbor-joining phylogenetic analysis of near full length genome sequences. The phylogenetic tree shows the different HIV-1 groups, subtypes and sub-subtypes. Each of the internal branches defining a subtype or sub-subtype is supported by 100% of bootstraps.

Groups refer to the very distinctive HIV-1 lineages M, N and O. The vast majority of HIV-1 strains found worldwide and responsible for the pandemic, belong to just one of these lineages, group M (for Major). Group O seems to be endemic to Cameroon and neighboring countries in West Central Africa, but even there these viruses represent a minority of HIV-1 strains: their highest
Recombinant HIV Sequences

prevalence is 2–5% of HIV-1 positive samples (55, 71, 113). Group N (for New, or non-M, non-O) has only recently been identified, and is so far represented by a limited number of isolates from Cameroonian patients (7, 91).

Within group M, there is further phylogenetic structure, allowing the classification of HIV-1 M strains. Subtypes were proposed because most sequences of group M were found to fall into a limited number of discrete clades (51, 52). The subtypes are approximately equidistantly related and in order to be considered as a subtype, isolates should resemble each other, and no other existing subtype, across the entire genome. In this light, there are only nine subtypes of HIV-1 group M, (A–D, F–H, J and K). In the case of subtype G, there is some ambiguity about the origins of the accessory gene region, which is close to subtype A (13, 23, 83), however, most of the subtype G genome is phylogenetically distinct. Phylogenetic analyses of group O strains have not revealed the same substructure as found within the evolutionary tree of group M, and so this group has not been classified into subtypes.

Within some subtypes, further phylogenetic structure can be identified, leading to a classification into subclades. Subtype F is subdivided into 2 subclades, F1 and F2 (100) and it is clear that subtypes B and D would be better considered as subclades of a single subtype, but for historical reasons it is difficult to change these designations. Also within subtype A, sub-subtype A2 strains have been recently described (26).

Subsequent to the designation of group M subtypes, it was realized that certain isolates clustered with different subtypes in different regions of their genomes when phylogenetic analyses were performed (81). Some of these mosaic HIV-1 genomes have been identified in several, apparently unlinked, individuals and play a major role in the global AIDS epidemic and are now designated as “Circulating Recombinant Forms”, or CRFs (15). Members of a CRF should resemble each other over the entire genome, with similar breakpoints reflecting common ancestry from the same recombination event(s). There are currently several CRFs of HIV-1: under new nomenclature proposals, each will be designated by an identifying number, with letters indicating the subtypes involved (83). If the genome contains sequences originating from more than two subtypes, the letters will be replaced by “cpx”, denoting “complex”.

In order to define a new subtype, sub-subtype or CRF, representative strains must be identified in at least three individuals with no direct epidemiological linkage. Three near full-length genomic sequences are preferred, but two complete genomes in conjunction with partial sequences of a third strain are sufficient to designate a new subtype, sub-subtype or CRF (to define a CRF, the partial sequence(s) must also confirm the CRFs mosaic structure).

INTERSUBTYPE RECOMBINATION OF HIV-1 GROUP M STRAINS

Overview of the actually known CRFs

The different CRFs, actually known are summarized in Table 1 and Figure 2 shows the complex mosaic genomic structure for each of them.

Table 1 Summary of the defined Circulating Recombinant Forms (CRF) of HIV-1 group M

<table>
<thead>
<tr>
<th>name</th>
<th>subtypes involved</th>
<th>geographic distribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>CRF01-AE</td>
<td>A, E</td>
<td>predominant in Southeast Asia, sporadic in Central Africa</td>
</tr>
<tr>
<td>CRF02-AG</td>
<td>A,G</td>
<td>predominant in West and West Central Africa</td>
</tr>
<tr>
<td>CRF03-AB</td>
<td>A, B</td>
<td>Russia (Kaliningrad)</td>
</tr>
<tr>
<td>CRF04-cpx</td>
<td>A, G, H, K, U</td>
<td>Greece, Cyprus</td>
</tr>
<tr>
<td>CRF05-DF</td>
<td>D, F</td>
<td>Democratic Republic of Congo</td>
</tr>
<tr>
<td>CRF07-BC</td>
<td>B, C</td>
<td>northwest China</td>
</tr>
<tr>
<td>CRF08-BC</td>
<td>B, C</td>
<td>southeast China</td>
</tr>
<tr>
<td>CRF09-cpx</td>
<td>unpublished</td>
<td>Senegal, US</td>
</tr>
<tr>
<td>CRF10-CD</td>
<td>C, D</td>
<td>Tanzania</td>
</tr>
<tr>
<td>CRF11-cpx</td>
<td>A, E, G, J</td>
<td>Central Africa(Cameroon, Central African Republic, Gabon)</td>
</tr>
</tbody>
</table>

CRF01-AE

All known representatives of what was initially described as subtype E appear in fact to be recombinants of subtypes A and E (12, 22), and are now designated CRF01-AE (83). Subtype E was first designated on the basis of the distinct phylogenetic position of these viruses in env trees, the only non-subtype A sequences are found within (most of) the env gene, parts of vif, vpr and nef, and the LTR. A full length non-recombinant subtype E sequence has not yet been described and the absence of one of the “parental” lineages
Recombinant HIV Sequences

Figure 2: Schematic representation of the complex mosaic genomic structure from the actually published circulating recombinant forms (CRFs).
Recombinant HIV Sequences

leads to difficulties in formally proving the recombinant nature of these viruses (3). CRF01-AE viruses have been documented at low frequencies in several Central African countries, like Central African Republic, Cameroon and the Democratic republic of Congo (64, 66, 105), but they are responsible for the explosive epidemic in Southeast Asia, especially in Thailand from where these viruses have further spread to surrounding countries like Vietnam, Cambodia, Myanmar and China (42, 58, 59, 62, 75, 108).

CRF02-AG

The IbNg strain from Ibadan, Nigeria, was initially described as a divergent lineage within subtype A, based on gag and env sequences (34). However, after the determination of a full length sequence, IbNg was recognized as a complex mosaic of alternating subtype A and subtype G sequences (13). Since a number of similar viruses have been reported from countries in both West and East Africa, this clade is now designated as CRF02-AG. CRF02-AG is the predominant HIV-1 strain in West and West Central Africa, where they represent between 50% to 70% of the circulating strains (4, 56, 61, 67, 73, 98). These viruses have now also been introduced in Europe and to a minor extent in the US (18, 57).

CRF03-AB

An epidemic among intravenous drug users (IDUs) in Kaliningrad, Russia, involves viruses that are mosaics of subtypes A and B (9, 48), and so are termed CRF03-AB. subtype A and B strains from Ukrainian IDUs were shown to be the probable parental viruses of the Kaliningrad AB recombinant strain (49). The epidemic in Kaliningrad was explosive and large numbers of simultaneous infections with this particular HIV-1 variant occurred, offering a unique opportunity to study the virus in a setting where the epidemic was captured in the earliest stages of spreading.

CRF04-cpx

Isolate 94CY032 from Cyprus was designated as the prototype of subtype I based on gp120 sequences (43). However, full genome sequencing revealed this virus was a complex recombinant with A, G and a putative new subtype, I. Multiple breakpoints were observed between the distinct subtypes (25) and two similar viruses have been reported from epidemiologically unlinked individuals from Greece (65). Re-analysis with previously unavailable complete genome sequences revealed that some of the unknown regions were in fact subtype H or K, but still some regions could not be classified. Thus, subtype I was removed from the genetic classification system of HIV strains, and the “I” regions were relabelled as unclassified (U). These strains are now called CRF04-cpx, and their genome is comprised of subtype A, G, H, K and unknown fragments with multiple breakpoints (83, 88).

CRF05-DF

Two full-length strains were shown to be mosaics of subtypes F and D, CRF05-DF. These epidemiologically unrelated F/D sequences showed similar chimeric structure and partial sequences from three additional unlinked F/D recombinants confirmed this. Genetic distances in the phylogenetic trees suggest that the recombination event leading to the putative CRF occurred relatively long ago. Furthermore, all five F/D recombinants are linked to the Democratic Republic of Congo, suggesting that the original recombination event took place in central Africa (47).

CRF06-cpx

Two near-full-length genomes of similar complex mosaic viruses, containing fragments of (at least) subtypes A, G and J, have recently been described in patients from Burkina Faso (BFP90) and Mali (95ML84) (60, 68). Phylogenetic and recombinant analysis from two additional full-length genome sequences from epidemiologically unlinked individuals, one from Senegal (97SE-1078) and one from Mali (95ML-127), had a similar mosaic structure and confirmed that the previously described strains, BFP-90 and 95ML-84, represent a new CRF of HIV-1, designated as CRF06-cpx, since 4 different subtypes were involved in the mosaic genome structure. This new CRF was composed of successive fragments of subtype A, G, K and J. The fragment in the pol gene that was initially characterized as unknown in the BFP-90 strain and subsequently as subtype I in the 95ML-84 strain, was now clearly identified as subtype K. CRF06-cpx circulates in Senegal, Mali, Burkina Faso, Ivory Coast, Niger and Nigeria, although the exact prevalences remain to be determined (Toure Kane C, Montavon C, Nkengasong J, Saidou M, Peeters M, personal communication). Importantly, this new variant was also introduced in other continents, Europe (France) and Australia showing that these viruses are present not only locally but also globally. (68, Montavon C, Peeters M, personal communication).

CRF07-BC and CRF08-BC

Two different BC recombinants have been detected in intravenous drug users (IDU) in China. CRF07-BC, with the 97CN-54 prototype strain, are
isolated among IDUs in the northwestern part of China (95). CRF08-BC, prototype 97CN-6F, is documented in IDU in Guangxi, southern China neighboring Myanmar (75). CRF08-BC strains are mostly subtype C with portions of the capsid and reverse transcriptase genes from subtype B. Whereas the breakpoint in p24/p17 and the RT gene overlap, CRF07-BC strains have additional breakpoints in the p7/p6 genes, the vpr/vpu, and in the 3' portion of nef. The two parental B'-Thai and C subtypes have been reported earlier to co-circulate among IDUs in southwestern China, therefore clearly representing a potential reservoir for recombination (29, 53, 111, 112). Two different routes of BC recombinants spread throughout China, suggesting different founder effects in the Chinese IDU population (75).

**CRF09-cpx, CRF10-CD and CRF11-cpx**

CRF09-cpx has been described in Senegal and a US military seroconverter (57). CRF10-CD was recently described in a cohort studying mother to child transmission of HIV-1 in Dar-es-Salaam, Tanzania (44). In this country, subtypes A, C and D cocirculate in equal proportions and many samples with discordant subtype designations between 2 or more genomic regions have already been documented (33, 79, 80). CRF11-cpx, involving subtypes A, G, J and E is observed in Cameroon and the Central African Republic (CAR). These strains were also described in individuals from Europe, but infected with HIV overseas, more precisely in Djibouti, CAR and even in French Guyana. A recently reported A, G, E, ? strain, isolated from a patient in Greece, seems after reanalysis of the genome, to have a similar mosaic structure as CRF11-cpx. (70, 101, Montavon C, Peeters M, personnel communication)

The majority of CRFs have been documented in local epidemics only. This is the case for CRF03-AB, CRF04-cpx, CFR05-DF, CRF07-BC, CRF08-BC, and CRF10-CD. Some are spreading into different countries, but actually their prevalence seems to be low, like CRF06-cpx, CRF09-cpx and CRF11-cpx. CRF01-AE and CRF02-AG, however, account for large numbers of HIV-1 infections worldwide, and play a major role in the global epidemic, in southeast Asia and Africa respectively, and they are also introduced to other continents.

**Unique recombinants**

In addition to all this circulating recombinants, full-length genome sequences from many more unique recombinants have been described. Several AC and AD recombinants have been described in Eastern Africa (14, 45, 57, 87), where these 3 subtypes co-circulate. B/F recombinants have been found in Brazil and Argentina, where subtypes B and F are both common (54, 84, 85). Subtypes A and C co-circulate in India, and A/C recombinants are present (37, 50). Recently, a B/CRF01-AE was observed in Thailand were subtype B and CRF01-AE were initially introduced (99). Various other complex recombinants including even small or large fragments from unclassified sequences have been reported from Africa where all subtypes cocirculate. In addition, some of the first African HIV-1 isolates to be characterized, MAL and Z321 (obtained from a stored plasma sample obtained in 1976 in a rural area in the north of the DRC (93)), have been identified as complex recombinants, A/D/? and A/G/? respectively (2, 16). Some of the unknown fragments can be present in two different recombinants. For example, the 97CD-KTB49 strain from DRC is a complex recombinant involving subtypes A, E, G, H, J, K and unknown fragments in the vif-vpr region (106). Additional analyses confirmed also that the vif-vpr fragments consists of 2 different unknown fragments with the 5' end of the vif gene corresponding to the unknown fragment of the Z321 strain and the 3' end of the vif and the vpr fragment corresponding to the unknown fragment common between Z321 and CRF04-cpx. Another example is the NOGIL3 virus from a family in Norway (40), where some of the unclassified fragments correspond to unknown fragments observed in the Mal virus. Figure 3 illustrates the complex genomic structure of these viruses mainly from Central African origin. The presence of recombinant viruses early in the AIDS epidemic and the complexity as well as the numerous unclassified fragments found in recombinants that actually circulate in Central Africa, confirms that HIV was already present for awhile in this region of Africa.

**RECOMBINATION BETWEEN HIV GROUPS**

It was initially suspected that homologous recombination between group M and group O viruses may not be possible because of their high degree of divergence. Two recent reports however have documented intergroup recombinants in two different patients from Cameroon (72, 97). M/O mosaic viruses can replicate well in vivo and in vitro, and can even become the predominant variant within the patient's viral population (72). Recombination between such divergent strains could contribute substantially to the emergence of new HIV-1 variants, and would have important implications both for diagnosis by serological and molecular tests, and for treatment.

The phylogenetic position of YBF-30 and YBF-106, the only two group N representatives for which full genome sequences are so far available, depends of the gene studied. Using sequences from the 5' half of the genome, group N forms an independent lineage most closely related to, but still distant from,
Recombinant HIV Sequences

Figure 3: Schematic representation of the complex mosaic genomic structure from unique recombinant HIV-1 viruses from Central African origin. This figure illustrates that some unclassified fragments can be present in different recombinant viruses.
group M. In contrast, with sequences from the 3' half of the genome, group N viruses cluster more closely with a chimpanzee virus (SIVcpzUS) (17, 25). These data suggest that group N viruses are the result of a recombination event between an SIVcpz like and an HIV-1 like virus. This observation offers further substantiation of a chimpanzee-human zoonosis.

These observations also open the hypothesis that distant SIVs and HIVs can potentially recombine, particularly in individuals who are exposed to SIV by cross-species transmission. By this means novel SIV sequences may be introduced more efficiently into the human population.

While dual infections with HIV-1 and HIV-2 have frequently been reported in regions where both viruses circulate (41), as yet no recombinants between them have been described. In this case, the level of genetic divergence may be too high for successful recombination, although its possibility cannot be entirely excluded.

METHODS TO IDENTIFY RECOMBINANTS

Recombinants can only be detected if different parts of the genome are genetically characterized, either by sequencing or by more simplified subtyping techniques such as HMA in env and gag (19, 30). Nevertheless, sequencing remains the most accurate approach to identify HIV-1 variants, especially recombinants or CRFs. Even only partial gag and/or env sequences give more precise information than HMA with regard to the presence of subclades or recombinant viruses (Figure 4). CRFs can form subclusters within a certain subtype. For example in phylogenetic trees, CRF02-AG strains form a different subcluster among subtype A in env and gag, whereas env HMA cannot discriminate CRF02-AG strains from subtype A. Also, CRF06-cpx strains form a separate subcluster within subtype G in the envelope, and a separate subcluster within subtype A in the gag region. Subtype A sequences from the CRF01-AE strains, also form a separate subcluster, see Figure 4.

Only full-length sequencing can determine the exact pattern of mosaicism within an isolate that is recombinant. A variety of complementary approaches have been developed to identify sequences that are recombinants, and to map the positions of breakpoints within mosaic sequences. Because the different subtypes of HIV-1 have been well defined, potential intersubtype recombinants can be analyzed in a more-or-less automated fashion. Using moving window analysis, diversity or similarity plots (23, 50) can be used to display the extent of difference or similarity of a new sequence to representatives of other subtypes. The Recombinant Identification Program (RIP) (89) uses distance measures to assign subtypes to regions within the new sequence. Bootscanning assesses the strength of bootstrap support for the phylogenetic placement of the new sequence with any subtype representative. For bootscanning, neighbor joining trees for windows moving along the alignment are done and the bootstrap values for the studied sequences are plotted at the midpoint of each window resulting in bootstrap plots (86). Fine-scale mapping of recombination breakpoints has been performed using informative site analysis (23, 81). Much of the software used to perform these analyses is freely available from the authors, or can be accessed on-line. Summaries of and links to the different programs are available on the Los Alamos HIV Database website (http://hiv-web.lanl.gov/) and David Robertson's site at http://grinch.zoo.ox.ac.uk/RAP_links.html.

WORLDWIDE DISTRIBUTION OF HIV-1 VARIANTS

Subtype designations have been powerful molecular epidemiological markers to track the course of the HIV-1 pandemic. It seems clear that the various subtypes, subclades within subtypes, and CRFs have been generated by epidemiological accidents. Figure 5 shows the geographic distribution of HIV-1 subtypes and CRFs. The predominant viral forms in the global epidemic are subtypes A and C, followed by subtype B and the recombinants CRF01-AE and CRF02-AG (57, 74, 110). The greatest genetic diversity of HIV-1 has been found in Africa, especially Central Africa. Overall subtypes A and C and CRF02-AG are most common, but all groups and subtypes are found, consistent with this continent being the source of the epidemic (105). In South and East Africa subtype C predominates (31, 36, 104). In West and West Central Africa, the majority of viruses are CRF02-AG (61). In North America, Europe and Australia, subtype B is by far the most common. However, various other group M subtypes, and even group O viruses, have been reported in the US (5, 10, 11, 78, 109) and several European countries (1, 30, 46, 90) and there the unusual subtypes even seem to be increasing (8, 20, 92). In South America, subtype B predominates, but subtypes E and C are also found (38, 54, 84). Different subtypes circulate in Asia, subtype C predominates in India and CRF01-AE is predominant in southeast Asia.

The exact prevalence of recombinant strains is not well known, since few systematic studies have been conducted to address this problem. Preliminary data show that for example in Africa, the proportions of discordant gag/env samples can vary from less than 10% to up to 40% according to the countries or regions studied (56, 61, 73, 79, 80, 98, 105). The subtypes involved in these discordant samples, depend on the subtypes that co-circulate in a certain
Recombinant HIV Sequences

Figure 4: Evolutionary relationships of the Circulating Recombinant Forms, CRF01-AE, CRF02-AG, CRF06-cpx and CRF11-cpx in different regions of the genome, based on neighbor joining phylogenetic trees of partial gag sequences (p24) and partial env sequences (V3-V5).

Env (V3-V5)  Gag (p24)
Figure 5: Geographic distribution of HIV-1 subtypes and CRFs.
region, for instance, in Nigeria only subtypes A and G co-circulate and are the only subtypes involved in the 37% discordant samples (56, 73). As expected, given the presence of numerous co-circulating subtypes, a wide variety of recombinants has been reported in DRC and all subtypes are involved in the 29% discordant samples (106).

The global distribution of different forms of HIV-1 is a dynamic process. As more HIV-1 variants inevitably intermix in different parts of the world the likelihood of generating new recombinant viruses will increase. The pattern of mosaicism will become even more complex, since recombination involving viruses that are already recombinant will occur. Mosaics involving CFR02_AG have already been observed in various African countries (39, 61, 72). Continued monitoring is necessary to determine the future role of non-subtype B viruses in North America and Europe, and to chart the emergence of new predominant subtypes and CRFs around the world.

IMPLICATIONS OF RECOMBINATION

Recombinant viruses may have some advantages over the parental strain, including eventual modifications in tropism and replication efficiency (fitness). Several studies have found that, under the selective pressure imposed by antiretroviral drugs, recombination between strains with different drug sensitivities occurred, resulting in new HIV-1 variants with dual drug resistance (63). In vitro experiments with feline and murine retroviruses have demonstrated that mixed infections can generate recombinant viruses with altered tissue tropism, pathogenicity, and host range, or with changes in antigenic epitopes (27, 102). Finally, recombination also has important implications for vaccine strategies based on live-attenuated viruses, since these could recombine with infecting strains, even if the two are quite divergent.

The discovery of large numbers of recombinant viruses clearly implies that co-infection with divergent HIV-1 strains is not as rare as once thought. Indeed, dual infections with different subtypes have been reported in regions where multiple variants co-circulate (6, 38, 72, 77, 96, 114). It remains to be determined when superinfection can occur during the course of HIV infection. In macaques, it has recently been shown that superinfection with divergent strains of HIV-2 is possible only during certain periods when antibodies are not yet efficiently expressed (69). In contrast, in a chimpanzee, superinfection with a CRF01-AE virus 32 weeks after experimental infection with a subtype B strain led to a dual infection with the rapid appearance of recombinant viruses (21). The latter result suggests that superinfection is not restricted to the early phase of infection, and implies that the humoral and cellular antibody response are not efficient against divergent strains.

CONCLUSIONS AND PERSPECTIVES

The geographical distribution of subtypes is a dynamic and unpredictable process and intermixing of HIV-1 variants is inevitable. Recombinant viruses contribute already substantially to the global pandemic, and the likelihood of generating recombinant viruses will only continue to increase as the different HIV-1 subtypes spread to all continents, and even recombinant viruses will recombine. The proportion of recombinant viruses will depend on the prevalence rates of different subtypes, the probability that certain population groups acquire multiple infections and transmit their viruses further, and the fitness of any mosaic viruses generated. However, the frequency of recombinant viruses is almost certain to increase; recombination, once it has occurred, cannot be undone. In future molecular epidemiologic studies, pure subtypes and CRFs have to be monitored. More studies are needed to understand the role and the implications of recombinant viruses in the global HIV evolution. It is important to study in more detail the impact of viral recombination on viral properties, since recombination may introduce genetic and biological consequences that are far greater than those resulting from the steady accumulation of single mutations. In order to develop an efficient vaccine, it remains to be determined when superinfection can occur during the course of HIV infection, and to what extend humoral and cellular immune response are efficient against divergent strains.

REFERENCES


[16] Choi DJ, Dube S, Spicer TP, Slade HB, Jensen FC, Poiesz BJ. HIV type 1 isolate Z321, the strain used to make a therapeutic HIV type 1 immunogen, is intersubtype recombinant. *AIDS Res Hum Retroviruses* 1997, **13**:357–361.


Recombinant HIV Sequences


[57] McCutchan F. Understanding the genetic diversity of HIV. *AIDS* 2000, 14 (suppl), S31-S44.

Recombinant HIV Sequences


Recombinant HIV Sequences


Recombinant HIV Sequences


