Impact of genetic diversity of HIV-1 on diagnosis, antiretroviral therapy & vaccine development

Renu B. Lal, Sekhar Chakrabarti* & Chunfu Yang

Divisions of HIV/AIDS Prevention, National Center for HIV, STD & TB Prevention, Centers for Disease Control & Prevention, Atlanta, USA & *National Institute of Cholera & Enteric Diseases (ICMR), Kolkata, India

Accepted February 17, 2005

HIV-1 strains have diversified extensively through mutation and recombination since their initial transmission to human beings many decades ago in central Africa. The high error rate of HIV reverse transcriptase combined with the estimated in vivo HIV-1 replication rate of ten billion new virions each day leads to extraordinary genetic diversity of HIV. Twenty seven circulating genetic forms of the HIV-1 group M are presently recognized, including 11 subtypes and sub-subtypes, and 16 circulating recombinant forms (CRF). Genotypic analyses have provided a better understanding of the molecular diversity of HIV-1, enabling the detection of emerging HIV-1 variants and improving the tracking of the epidemic worldwide. The rapid evolution of HIV within infected hosts contributes significantly to the elusiveness of this pathogen from host antiviral responses. The complex nature of HIV envelope glycoprotein that is inherently resistant to neutralization, the selective infection, progressive destruction and impaired regeneration of CD4+ T helper cells, generation of cytotoxic T lymphocyte (CTL) escape mutants, together with high genetic diversity with continually evolving HIV variants worldwide, makes design of an effective vaccine a formidable task. Given the rapidity and unpredictability with which HIV-1 genetic forms may propagate in future, a vaccine protective against all major HIV-1 circulating genetic forms is desirable, which could require multivalent formulations. Understanding the kinetics and directions of this continuing adaptation and its impact on viral fitness, immunogenicity and pathogenicity are crucial to the successful design of effective HIV vaccines. In this review, we focus on extensive diversity of HIV-1, emergence of recombinant forms and their impact on diagnosis, antiretroviral therapy, disease progression, transmission, and vaccine development.

Human immunodeficiency viruses (HIV-1 and HIV-2) are the etiologic agents for AIDS in humans1-3. HIV-1 has spread to most parts of the world, while HIV-2 has remained largely restricted to West Africa1-3. AIDS was first recognized in the 1980s2, and is the leading cause of death in many developing countries. UNAIDS estimates that 37.8 million individuals are infected with HIV, of which about 70 per cent live in sub-Saharan Africa and Asia4. Despite extensive preventive programmes worldwide, 4.8 million new infections occurred in 20034. Genetic variation is inherent to all RNA viruses
but has been best characterized for HIV-1. The extensive heterogeneity observed in the worldwide epidemic of HIV-1 originates from the rapid viral turnover (10^{10} viral particles/day) in an HIV infected individual, and the high rate of incorrect nucleotide substitutions during HIV reverse transcription (10^{-4}/nt) in the absence of proof-reading mechanisms. It is clear that zoonotic transmission of lentiviruses plays an important role in the emergence of human retroviruses.

HIV-1 continually evolves and migrates through individual hosts, overcoming barriers to transmission, avoiding different immune responses, and resisting various antiretroviral regimens. While a vaccine (preventive or therapeutic) is the only hope to curtail the epidemic, the diversity of HIV-1 provides an extraordinary challenge in drug and vaccine development. In the present review, we will focus on extensive diversity of HIV-1, emergence of recombinant forms and their impact on diagnosis, antiretroviral therapy, disease progression, transmission, and vaccine development.

Zoonotic transmission of the lentiviruses

HIV-1 and HIV-2 and the closely related simian immunodeficiency viruses (SIV) belong to the lentivirus subfamily of retroviruses. Humans are exposed to a plethora of primate lentiviruses through hunting and handling of primate bushmeat in central Africa. Both HIV-1 and HIV-2 are the result of zoonotic transmission of SIVcpz in chimpanzees (Pan troglodytes troglodytes) from West central Africa and SIVsm in sooty mangabeys (Cercocebus atys) from West Africa, respectively. Further, it is believed that at least two to three separate zoonotic jumps from chimpanzees into humans led to the disproportionate spread of HIV-1 groups M, O, and N. In addition, human have apparently picked up as many as seven lineages of viruses from sooty mangabeys resulting in HIV-2 subtypes A through G. The possibility of additional zoonotic transfers of primate lentiviruses from species other than chimpanzees and sooty mangabeys and the extraordinary impact that can result from such primate lentiviral zoonotic transmission events remains a continuum public health challenge.

Attempts to estimate the time of origin of HIV-1 using different methods of molecular clock analysis showed that the origin of HIV-1 group M radiation was in the 1930s. Korber et al. have estimated that the HIV-1 M group viruses last shared a common ancestor between 1915 and 1941. Salemi et al. used a different method and arrived at a similar date for the origin of the HIV-1 M group. They further calculated that the common ancestor of the HIV-1 M group and the SIVcpz isolated from Pan troglodytes troglodytes dated to the late 17th century with a 99 per cent confidence interval for a date between 1591 and 1761. The epidemic apparently went unnoticed in central sub-Saharan Africa for several decades. The initial diversification of group M may have occurred within or near the Democratic Republic of Congo, where the highest diversity of group M strains has been observed and the earliest known case of HIV-1 infection, dating from 1959, has been documented.

Genetic variants of HIV

One of the major characteristics of lentiviruses is their extensive genetic variability, which is the result of the high error rate, the recombinogenic properties of the reverse transcriptase enzyme and the fast turnover of virions in HIV infected individuals. HIV-1 has been divided into three groups, M, O and N. Within HIV-1 group M, which accounts for the majority of infections worldwide, most sequences fall into a limited number of discrete clades, allowing the classification of HIV-1 group M strains into subtypes and sub-subtypes. Currently nine subtypes of HIV-1 group M exist: A-D, F-H, J and K. Subtypes form clusters roughly equidistant from each other in phylogenetic trees. Based on such definition, it is clear that subtypes B and D would be better considered as sub-subtypes of a single subtype, however, it is difficult to change the nomenclature now. Within the subtypes A and F, separate sub-clusters are distinguished, designated...
sub-subtypes (or sub-clusters) A1 and A2, and F1 and F2, each pair of sub-subtypes being more closely related to each other than with other subtypes. Some subtypes share a geographic localization, whereas others appear to have a similar ancestry. Examples include subtype G viruses from Spain and Portugal, and subtype C viruses from areas as geographically diverse as India, Ethiopia and South Africa.

HIV-1 groups O and N, which are genetically very divergent from group M, represent less than 5 per cent of infections worldwide and have almost exclusively been detected in West Central Africa. Elsewhere in the world, group O viruses have been identified mainly from persons with epidemiological links to West Central Africa, mainly Cameroon and some neighboring countries. Further, HIV-1 group N infection has only been identified in Cameroon, a country endemic for HIV-1, with all the major groups in cocirculation. Given that the conditions favouring human exposure to chimpanzees have increased in Cameroon and other parts of Africa because of commercial logging and hunting practices, it is possible that additional new transmissions of this virus may have occurred.
Fig. 2. Complex mosaic genomic structures of 16 circulating recombinant forms (CRFs). Letters and colours represent the different subtypes of HIV-1 that comprise the CRFs. Data downloaded from HIV Sequence Database, Los Alamos, NM Source: http://www.hiv/anl.gov/content/hiv-db/CRFs/CRFs.html (accessed on November 2004).
HIV-1 intersubtype recombinants

In addition to the rapid accumulation of minor genotypic changes, different HIV-1 strains can also recombine at a high rate, generating large genetic alterations\textsuperscript{12,13,36,37}. Recombination requires the simultaneous infection of a cell with two different proviruses, allowing the encapsulation of one RNA transcript from each provirus into a heterozygous virion. After the subsequent infection of a new cell, the reverse transcriptase generates a newly synthesized retroviral DNA sequence that is recombinant between the two parental genomes\textsuperscript{16,36}. These mosaic viruses display discrete breakpoints between the genomic regions with different phylogenetic associations\textsuperscript{16,36}. The fact that large numbers of recombinant viruses exist clearly implies that co-infection with divergent HIV-1 strains is more frequent than previously thought\textsuperscript{12-15}. Indeed, dual infections with different subtypes have been reported in regions where multiple strains co-circulate\textsuperscript{48}. Recent studies have documented cases of superinfection with a virus from another subtype\textsuperscript{49,50} and provided evidence of recombination following superinfection\textsuperscript{51,52}. Dual infections with HIV-1 and HIV-2 have frequently been reported in regions where both viruses circulate, however, no recombinants between these two viruses have yet been described\textsuperscript{16}.

There are at least 16 different circulating recombinant forms (CRFs) identified to date\textsuperscript{16} (http://www.hiv.lanl.gov/content/hiv-db/CRFs/CRFs.html). By definition, CRFs should resemble each other over the entire genome, with similar breakpoints reflecting common ancestry from the same recombination event(s) (Fig.2A, 2B)\textsuperscript{39}. Presently, 16 CRFs of HIV-1 group M exits, each is designated by an identifying number, with letters indicating the subtypes involved, the letters are replaced by “cpx”, denoting “complex” if more than 2 subtypes are involved\textsuperscript{16,39}. A vast majority of the CRF genomes are highly complex intersubtype recombinants, with some being extremely stable and others having a patchy appearance due to multiple crossover points\textsuperscript{16}. For instance, CRF01_AE is clearly E in \textit{env}, U/A/E in the regulatory region, and A in \textit{gag} and \textit{pol}, while the CRF04_cpx has an extremely complex mosaic genome structure including subtypes A, G, H, K and U\textsuperscript{16}. To date, CRFs have been identified in nearly every region of the world where two or more subtypes co-circulate and may account for over 10 per cent of new HIV-1 infections\textsuperscript{12-16}.

CRF01_AE viruses are responsible for the explosive epidemic in Southeast Asia, especially in Thailand from where they have further spread to surrounding countries like Vietnam, Cambodia, Myanmar and China\textsuperscript{5} (Fig.3). The CRF01_AE have been documented at low frequencies in several Central African countries, like Central African Republic, Cameroon and the Democratic republic of Congo\textsuperscript{16}. CRF02_AG, complex mosaic of alternating subtype A and subtype G sequences, is the predominant HIV-1 strain in West and West Central Africa. These viruses have now also been introduced in Europe, the US and South America\textsuperscript{14,15,53-56}. CRF03_AB is the predominant CRF among intravenous drug users (IDUs) in Kaliningrad, Russia\textsuperscript{57,58}. CRF04_cpx, found in Cyprus, is a complex mosaic comprising subtypes A, G, H, K and unknown fragments with multiple breakpoints\textsuperscript{59,60}. All CRF05_DF identified to date are linked to the Democratic Republic of Congo, suggesting that the original recombination event took place in central Africa\textsuperscript{61}. CRF06_cpx is a complex mosaic composed of successive fragments of subtypes A, G, K and J\textsuperscript{62}. CRF06_cpx circulates in Senegal, Mali, Burkina Faso, Ivory Coast, Niger and Nigeria\textsuperscript{62,63}. Importantly, this new variant is also present in other continents, Europe (France) and Australia\textsuperscript{16}.

CRF07_BC and CRF08_BC are two different BC recombinants have been detected in IDUs in China\textsuperscript{64}. CRF07_BC is present in the northwestern part of China\textsuperscript{16,64-66}, whereas CRF08_BC is present in Guangxi, southern China neighboring Myanmar\textsuperscript{66}. 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 with CRF08, CRF07_BC strains have additional breakpoints in the p7/p6 genes, the vpr/vpu, and in the 3’ portion of nef\textsuperscript{16}. The two parental Thai-B and C subtypes have been reported earlier to co-circulate among IDUs in southwestern China, therefore clearly representing a potential
reservoir for recombination\textsuperscript{16,64}. CRF09\_cpx has been described in Senegal and a US military sero-convertor\textsuperscript{67}. CRF10\_CD was recently identified in Dar-es-Salaam, Tanzania\textsuperscript{68}. In this country, subtypes A, C and D co-circulate in equal proportions and many samples with discordant subtype designations between 2 or more genomic regions have already been documented\textsuperscript{68,69}. CRF11\_cpx, involving subtypes A, G, J and E is observed in Republic of Congo, Cameroon and the Central African Republic (CAR)\textsuperscript{70,71}. CRF12\_BF has been identified in Argentina and Uruguay with most of the genome sequence originated from subtype F and with small patches from subtype B\textsuperscript{72}.

CRF13\_cpx is a complex recombinants comprising subtypes A, G, J, and CRF01\_AE and was identified in Cameroon\textsuperscript{73}. Likewise, CRF14\_BG has been identified in Spain where the most prevalent subtype is B\textsuperscript{74}. It is interesting to note that some of the CRFs have recombined with other subtypes and given rise to new CRFs like CRF15\_01B found in Thailand\textsuperscript{75}.

\textbf{Fig. 3.} Geographical distribution of HIV-1 subtypes, and circulating recombinant forms (CRFs) in different parts of the world. Ten different epidemic patterns have been observed, as shown in different colours. (Data adopted from www.iavireport.org). Source: IAVI Report online (http://www.iavireport.org) accessed on November 2003.
The newest CRF, CRF16_A2D was identified in Kenya, South Korea and Argentina\textsuperscript{76}.

The pattern of mosaic recombinant viruses is becoming even more complex, since recombination involving viruses that are recombinant is already occurring\textsuperscript{16}. In some cases, unique recombinant form (URF) viruses seem to have originated by secondary recombination of a CRF\textsuperscript{73,77,78}. Mosaics involving CFR02_AG have been observed in various African countries, and some of the established CRFs, like CRF11 and CRF13, contain sequences that are derived from CRF01_AE\textsuperscript{77}. Recombinations between two CRFs have been described in studies from China\textsuperscript{77} and Niger\textsuperscript{78}, in the latter case, three full-length genome sequences revealed the presence of complex and diverse CRF02/CRF06 recombinants. Various other complex recombinants including even small or large fragments from unclassified sequences have been reported from Africa where all subtypes cocirculate\textsuperscript{16}.

The majority of CRFs have only been documented in local epidemics. This is the case for CRF03_AB, CRF04_cpx, CRF05_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. However, CRF01_AE and CRF02_AG account for large numbers of HIV-1 infections worldwide, and play a major role in the global epidemic, in southeast Asia and Africa\textsuperscript{4,38,40}, respectively, and they are also introduced to other continents\textsuperscript{53,55,79}. While the emergence of these recombinant viruses has important implications for both monitoring HIV-1 genetic diversity and developing effective vaccines, predicting which recombinant strains will lead to future expansion of the epidemic is difficult.

The subtypes of most HIV-1 strains can be determined by sequence analysis of any of the major regions of the genome. Intersubtype nucleotide sequence divergence may exceed 20, 15 and 25 per cent for gag, pol, and env, respectively\textsuperscript{80}. Subtyping from a single gene region should be done cautiously, as recombinant strains comprising two or more subtypes may be missed by this approach. 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. Only full-length sequencing can determine the exact pattern of mosaics within an isolate that is recombinant\textsuperscript{16,39}. A variety of complementary approaches have been developed to identify sequences that are recombinants, and to map the positions of breakpoints within mosaic sequences. The links to several websites in which different programmes for sequence analyses are available: http://hiv-web.lanl.gov/ at the Los Alamos HIV database website, http://grinch.zoo.ox.ac.uk/RAPlinks.html at David Robertson’s site, http://sray.med.som.jhmi.edu/RaySoft/ at Stuart Ray’s website and http://evolution.genetics.washington.edu/ at Joseph Felsenstein’s website.

**HIV intergroup recombinants**

Recombination between strains from distant lineages may contribute substantially to new HIV-1 strains and could have important consequences. Recombination between two highly divergent groups M and O of HIV-1 have been reported from Cameroon\textsuperscript{81,82}. Group 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\textsuperscript{81}. If recombinant inter-group viruses have a better fitness than the parental group O viruses, their prevalence may increase rapidly with consequences on serological and molecular diagnosis, and treatment since differences among susceptibilities to certain antiretroviral drugs have been observed in vitro\textsuperscript{83}.

Likewise, detailed phylogenetic analysis of two group N viruses have suggested that group N viruses are the result of a recombination event between an SIVcpz like and an HIV-1 like virus\textsuperscript{84}. Using sequences from the 5’end of the genome, group N forms an independent lineage most closely related to, but still distant from, group M, whereas sequences from the 3’ end of the genome cluster more closely with a chimpanzee virus (SIVcpzUS)\textsuperscript{16,84}. This observation offers further substantiation of a chimpanzee-human zoonosis\textsuperscript{7-9}. 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\textsuperscript{18,19}. The HIV-1 group M pandemic provides compelling evidence for the overwhelming impact that can result from even a single event of primate lentiviral zoonotic transmission\textsuperscript{7}.

**HIV-2 strains**

Both the epidemiology and the molecular phylogenies of HIV-2 viruses are less well understood than those of the HIV-1 M group\textsuperscript{12,17,26,85,86}. The diverse HIV-2 lineages are thought to have arisen in sooty mangabeys, with several cross-species transfers from sooty mangabeys into humans, one for each subtype of HIV-2\textsuperscript{17,27} (Fig.1). Thus far, seven subtypes (A-G) of HIV-2 have been described, with subtypes A and B being the predominant strains\textsuperscript{16,27}. The nucleotide and amino acid sequence diversity of HIV-2 viruses is greater than the diversity of the HIV-1 M group viruses, but roughly equivalent to the diversity within the HIV-1/SIVcpz clade (Fig.1). The subtypes of HIV-2 are thus analogous to the groups (M, N and O) of HIV-1, both in terms of sequence diversity and in terms of the cross-species transfer events thought to have created them\textsuperscript{17,25-27}.

Subtype A is most frequent in western part of West Africa (Senegal and Guinea-Bissau) and subtype B predominates in Ivory Coast\textsuperscript{3,26,27,86-88}. The other subtypes have been documented only in few individuals, with subtypes C, D, E and F found in rural areas in Sierra Leone and Liberia and subtype G from Ivory Coast\textsuperscript{3,27,86}. While dual infections with HIV-1 and HIV-2 have been frequently reported in regions where both viruses circulate\textsuperscript{16,85,86}, 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.

**Worldwide distribution of HIV-1 variants**

Divergent inter-patient HIV-1 evolution coupled with new introductions into susceptible human populations has lead to the current HIV-1 diversification from the original zoonotic transmission in central Africa\textsuperscript{7-9,12,13}. Although a founder effect results in further HIV-1 spread and divergent evolution in different regions in the world\textsuperscript{10,12,89}, the phylogenies of most non-African HIV-1 strains can be traced to central African isolates\textsuperscript{14,15,17,30}. Increased fitness has likely played a key role in the predominance and extreme variation of HIV-1 group M over group O or N isolates\textsuperscript{8,10,90}.

Genotypic analysis has provided a better understanding of the molecular epidemiology of HIV-1, enabling the detection of emerging HIV-1 variants and improving the tracking of the epidemic worldwide\textsuperscript{12-16}. 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\textsuperscript{38,40}. The map (Fig.3) summarizes current understanding of the global distribution of HIV-1 strains\textsuperscript{5}. Ten different epidemic patterns have been observed, as indicated by the different colours in the map which gives an overview of HIV-1 subtypes and recombinants in any given location. In the Americas, Western Europe and Australia, subtype B predominates, but in eastern South America, there is a substantial proportion of BF recombinants in addition to subtype B. In eastern Europe, subtypes A, B, and AB recombinant strains dominate the epidemic. Three different patterns have been observed in Asia: subtype C (India), a mixture of B, C, and BC Recombinants (China), and a mixture of subtype B and CRF01_AE (Thailand). Africa shows the greatest diversity. Subtype C dominates the South and East Africa, except for significant foci of subtypes A and D, as shown. West and West Central Africa harbour mainly CRF02_AG, alongside a complex array of other recombinants each present at low frequency. The most complex epidemic is in Central Africa, where rare subtypes and a wide variety of recombinant forms circulate without any discernible predominant strains\textsuperscript{30-33}. There is paucity of subtype information in Northern Africa, the Middle East, and Central Asia (gray).

**HIV subtypes in India**

Molecular epidemiologic studies carried out in different parts of India have indicated that HIV-1 subtype C is the most prevalent subtype in India\textsuperscript{91-101} (Fig.4). Few HIV-2 infections have been identified, almost all of these represent infections with HIV-2 subtype A\textsuperscript{102,103}. The distribution of the different HIV-
**Fig. 4.** HIV-1 subtype distribution, with predominance of subtype C in India. Note emergence of Thai subtype B and dual infections with subtype C and Thai subtype B.
1 group M variants in India based on published studies, which often used different molecular techniques including the heteroduplex mobility assay (HMA), and partial or full-length genome sequencing, are summarized in Table I. Initial analysis of small fragments of V3-V5 envelope region of subtype C sequences from India, along with sequences from other countries with predominance of subtype C, has revealed that Indian C sequences form a phylogenetically distinct lineage within subtype C, designated as C3 (Indian). Additionally while majority of subtype C infections reveal maximum homology to the C3-Indian reference strain (IND868), almost a third of the infections show homology to C2-Zambian strain (ZM18) and a minor percentage shows close resemblance to C1-Malawi strain (MA959)\(^9\)\(^\text{1},\)\(^\text{97}\). Overall, C3 lineage sequences are more closely related to each other (10% divergence) than to subtype C sequences from Botswana, Burundi, South Africa, Tanzania, and Zimbabwe (range, 15-21%). The prevalence of three positions identified as signature amino acid substitution sites for C3 subcluster sequences (K340E, K350A, and G429E) are high among Indian sequences (56-87%) than non-Indian Cs (<22%)\(^9\)\(^\text{5},\)\(^\text{97},\)\(^\text{99}\). To further examine if Indian sequences formed a geographic cluster within subtype C, a larger envelope sequence analysis using neighbor-joining method was carried out for several subtype C sequences (Fig.5). The Indian sequences usually clustered as separate groups, along with a single sequence from Mynamar and China. A similar observation of distinct cluster was observed for sequences from South America (Brazil, Argentina, Uruguay). In contrast, sequences from south and east Africa do not form discrete cluster (Fig.5), indicating that African subtype C sequences are almost no closer to each other than they are to all other subtype C sequences. The distinct Indian cluster observed in subgenomic envelope analysis has further been substantiated using full-length analysis. A recent study which analysed 40 full-length sequences of subtype C available in the data base in 2003, revealed that Indian sequences form a distinct geographic cluster supported by very high boot-strap values\(^10\)\(^\text{4}\). These data further support the notion that the Indian sequences are distinct from other subtype C, possibly the result of a single introduction followed by rapid spread\(^9\)\(^\text{7},\)\(^\text{10}\)\(^\text{5}\). Because of such geographic clustering, India’s current vaccine efforts have focussed on developing vaccine comprising Indian subtype C in hopes to minimize any impact of diversity on vaccine efficacy. Indeed, clinical trends on modified vaccinia Ankara (MVA) recombinant using Indian subtype C are developed jointly by ICMR and IAVI and Phase I clinical trials started in February, 2005.

Limited studies have provided evidence for non-C subtypes, including some unique recombinant forms, in different parts of India (Table I). The first A/C recombinant was detected in India after full length genomes of 6 seroconverter from an STD cohort in Pune\(^9\)\(^\text{2}\). Additional studies from various parts of the country have identified infection with subtype A, Thai B and CRF01_AE. More importantly, Thai B is recognized as the second major subtype circulating in northeastern regions\(^9\)\(^5,\)\(^9\)\(^9\). The classification of HIV strains is evolving and depends on the sensitivity of techniques used in molecular epidemiology. Because most of the studies in India were based on a single gene region analysis, or used the techniques that are not sensitive enough to discriminate CRFs from unique recombinants, the role played by CRFs may be underestimated in these studies. For example, many studies used HMA, which needs less sophisticated equipment but is not sensitive enough to discriminate between pure subtypes and CRFs. Systematic studies involving better sampling and uniform technologies are needed to better understand the evolving HIV-1 epidemic in the country, especially in high risk groups from northeastern region of India.

Dual infections with subtype C and Thai B have recently been observed in Manipur, a region in northeast India with very high prevalence of HIV-1 among IDUs\(^9\)\(^3,\)\(^9\)\(^9,\)\(^10\)\(^6\). In addition, recent unpublished data have revealed discordant subtypes in gag and env regions of 4 (14%) of the 28 IDU studied from Manipur (Chakrabarty, unpublished), suggesting the presence of potential C/Thai B recombinants. Studies are in progress to ascertain if these infections resemble the CRF07/08 BC recombinants identified in China\(^6\)\(^4-\)\(^6\)\(^6\) or represent unique recombinant forms. Overland heroin export routes have been associated with dual epidemics of injecting drug use and HIV infections in Asian countries\(^10\)\(^6\). Whether drug
Fig. 5. Phylogenetic analysis of envelope sequences of diverse HIV-1 subtype C. A neighbor-joining tree was generated and the stability of tree topography was tested by bootstrapping and only significant bootstrap values that define geographic clusters are shown. Two geographic clusters with high bootstrap values include Asia (IN: India; CN: China; MM: Mynamar) and South America (BR: Brazil; AR: Argentina; UR: Uruguay), whereas East and South African specimens did not form a cluster (ET: Ethiopia; ZA: South Africa; TZ: Tanzania; BW: Botswana; KE: Kenya; BI: Burundi; DJ: Dijoundi) are shown.
### Table I. Distribution of HIV-1/HIV-2 subtypes in India

<table>
<thead>
<tr>
<th>City/State</th>
<th>Risk group</th>
<th>No. tested</th>
<th>Gene region</th>
<th>Subtype</th>
<th>Subcluster</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pune</td>
<td>STD clinic attendee</td>
<td>46</td>
<td>V3V5 HMA&lt;sup&gt;2&lt;/sup&gt;</td>
<td>96% C</td>
<td>66% C3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2% A</td>
<td>34% C2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2% B</td>
<td></td>
</tr>
<tr>
<td>Pune, Delhi&lt;sup&gt;34&lt;/sup&gt;</td>
<td>Recent seroconverter</td>
<td>6</td>
<td>Full length sequencing</td>
<td>83% C</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>17% A/C</td>
<td></td>
</tr>
<tr>
<td>Pune, Delhi&lt;sup&gt;34&lt;/sup&gt;</td>
<td>1992-1996</td>
<td>28</td>
<td>LTR</td>
<td>100% C</td>
<td></td>
</tr>
<tr>
<td>New Mumbai&lt;sup&gt;36&lt;/sup&gt;</td>
<td>Heterosexual</td>
<td>21</td>
<td>V3/ Vpr</td>
<td>B and C</td>
<td></td>
</tr>
<tr>
<td>Mumbai&lt;sup&gt;101&lt;/sup&gt;</td>
<td>Drug naïve cases</td>
<td>128</td>
<td>Protease/RT</td>
<td>96% C</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2% A/C</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.8% A</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.8% AE</td>
<td></td>
</tr>
<tr>
<td>Delhi&lt;sup&gt;98&lt;/sup&gt;</td>
<td>Seropositive patients in AIIMS, Delhi</td>
<td>125</td>
<td>V3V5</td>
<td>78% C</td>
<td>68% C3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>11% B’</td>
<td>29% C2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2.4% A</td>
<td>3% C1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2% E</td>
<td></td>
</tr>
<tr>
<td>Manipur&lt;sup&gt;93&lt;/sup&gt;</td>
<td>IDU</td>
<td>233</td>
<td>Peptide assay</td>
<td>71% C</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>29% B’</td>
<td></td>
</tr>
<tr>
<td>Kolkata&lt;sup&gt;95&lt;/sup&gt;</td>
<td>FSW</td>
<td>40</td>
<td>V3V5 HMA</td>
<td>95% C</td>
<td>50% C3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>5% Untypeable</td>
<td>29% C2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3% C1</td>
</tr>
<tr>
<td>Kolkata&lt;sup&gt;97&lt;/sup&gt;</td>
<td>FSW</td>
<td>21</td>
<td>V3V5</td>
<td>95% C</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>5% A</td>
<td></td>
</tr>
<tr>
<td>Kolkata&lt;sup&gt;99&lt;/sup&gt;</td>
<td>FSW</td>
<td>54</td>
<td>V3V4</td>
<td>92% C</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>8% Untypable</td>
<td>50% C3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>13% C2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2% C1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>27% C-Untypable</td>
</tr>
<tr>
<td>Manipur&lt;sup&gt;99&lt;/sup&gt;</td>
<td>IDU</td>
<td>25</td>
<td>V3V4</td>
<td>68% C</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>20% ThaiB</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>12% Dual (ThaiB/C)</td>
</tr>
<tr>
<td>South India&lt;sup&gt;100&lt;/sup&gt;</td>
<td>Seropositive from Hospital, Pvt. Clinic</td>
<td>256</td>
<td>LTR</td>
<td>98.8% C</td>
<td>0.4% A</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.4% B</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.4% B/C</td>
<td></td>
</tr>
<tr>
<td>South India&lt;sup&gt;102&lt;/sup&gt;</td>
<td>Seropositive</td>
<td>18</td>
<td>V3 region</td>
<td>HIV-2 subtype</td>
<td></td>
</tr>
<tr>
<td>Kolkata&lt;sup&gt;103&lt;/sup&gt;</td>
<td>Patients attending clinic</td>
<td></td>
<td></td>
<td>A</td>
<td></td>
</tr>
</tbody>
</table>

STD, Sexually transmitted disease; HMA, heteroduplex mobility assay; IDU, intravenous drug user; FSW, female sex worker; RT, reverse transcriptase.
Superscript numerals represent reference numbers.
trafficking across the Burma-India border to Manipur, with predominant subtype C, B and CRF01_AE, will result in new waves of recombinant infections in the bordering cities needs to be analyzed extensively. The recent findings of 40 per cent prevalence in high-risk groups in Kolkata, together with an estimated annual incidence of 7 per cent, provide a fertile enunciation for any recombinants that have better replicative or fitness advantage over the existing strains to emerge as the predominant strain in this area.107 Emerging trends of unique recombinants in neighbouring northeastern countries, including Bangladesh,108 Mynamar,109 Thailand,110 and China,64,66 can significantly impact the landscape of the epidemic, including vaccine design strategies in India.65,111,112

**Impact of HIV variability on diagnosis**

HIV diversity has enormous impact on diagnostics, as commercial assays must detect all known variants of HIV-1.113,114 Serologic assays are based on detection of antibodies to structural proteins, p24 and gp41, which are conserved among most HIV-1 isolates.80 Earlier studies had shown that persons infected with highly divergent HIV-1 group O, failed to be diagnosed accurately by some serologic tests due to gp41 variation.45,113 This problem was overcome by addition of group O lysate/peptide in the standard antigen mixture. Although most of current enzyme immuno assays (EIA) and rapid tests are sensitive and specific for diagnosing persons with chronically established infections with HIV-1 group M subtypes, the, some difficulties still remain in diagnosing persons with recent infections with non-B subtypes primarily because antigens used for the assays were based on HIV-1 subtype B strains.118 Continued analysis of commercial assays is needed to ensure that newly emerging variants do not escape detection by these tests.113,119

While genetic variability has had minimal impact on serologic detection of HIV-1 (due to amino acid conservation of the immunodominant epitopes on structural proteins), molecular diagnostic assays have been more challenging.113,114 Several generic primers that can identify all subtypes of HIV-1, groups M, N and O and SIVcpz have been developed as tools for detection and confirmation of HIV/SIVcpz infections worldwide.120 These primers provide a tool that not only detects infection with divergent viruses, but also allows subsequent phylogenetic analysis to identify subtypes for molecular epidemiologic studies.120,121 The utility of molecular detection of plasma viremia has increased considerably in clinical settings, especially for monitoring patients on antiretroviral (ARV) therapy. Four commercial viral-load assays are available: Amplicor HIV-1 Monitor v1.5 (Roche Molecular Diagnostics), nucleic-acid-sequence-based amplification (NASBA, Organon Teknika), HIV-1 Quantiplex branched DNA (bDNA, Chiron Diagnostics), and the LCx HIV-RNA quantitative assay (LCx, Abbott). The ability of these assays to quantify viral load from non-B subtype infections has been examined in several studies.122-127 Although, much progress has been made to improve the sensitivity of nucleic acid-based assays to quantify viral load across all subtypes of HIV-1, some RNA viral load assays still produce erroneous results.128-130 Thus continued monitoring of viral load assays against multiple subtypes of HIV-1 is needed to ensure accurate measurement of viral load in diverse populations infected with various subtypes of HIV-1.

**Impact of HIV-1 subtypes on antiretroviral therapy**

Antiretroviral drug resistance is the major obstacle for clinical management of ARV therapy.131-133 Current antiretroviral drugs comprise at least 16 drugs targeted against two pol gene enzymes, the reverse transcriptase (RT) and protease.133 The development of drug resistance to the nucleoside RT inhibitors (NRTIs), non-nucleoside RT inhibitors (NNRTIs) and protease inhibitors (PIs), is both a cause and a result of virologic treatment failure with incomplete virus suppression.133,134 Our understanding of drug resistance is largely limited to developed countries with predominant HIV-1 subtype B infections.132-136 However, with introduction of ARV in other parts of the world, emerging data indicate that viral subtype may influence the effectiveness of antiretroviral treatment and that pre-existing mutations could reduce the effectiveness of ARV.137-139 In the absence of any drug exposure, RT and protease sequences from B and non-B HIV-1 are polymorphic among about
40 per cent of the first 240 RT amino acids and 30 per cent of the 99 protease amino acids. The genetic baseline diversity of protease and RT from non-B subtype HIV-1 infected persons not exposed to antiretroviral therapy is summarized in Table II. Some of these amino acid substitutions occur at high rates in non-subtype B viruses at positions associated with drug resistance in subtype B.

The HIV-1 protease is a dimeric and symmetric molecule made up of two 99-residue polypeptide chains. The substrate cleft of the protease is formed by residues at positions 8, 23, 25-27, 28-30, 32, 47-52 and 53, 80, 82 and 84. These predicted amino acid residues appear to be highly conserved across subtypes (with the exception of V82I mutation in subtype G) in drug naïve persons. There are a number of amino acid substitutions that occur at high rates in certain non-subtype B viruses at positions associated with subtype B drug resistance (Table II). The V82I mutation occurs in about 1 per cent of untreated individuals infected with subtype B, it has been identified at 6 per cent in subtype C, 9 per cent in subtype F, and is the consensus (>50%) in subtype G isolates from untreated persons. This mutation alone confers minimal resistance to the available PIs in subtype B, however amino acid changes at V82 resulting in V82A, F, S or T are associated with high level resistance to most PIs. In subtypes F and G, the M46L or I substitution, which leads to resistance to most of the PIs in subtype B, has been identified in 4 and 7 per cent of sequences from untreated individuals, respectively.

Sequence comparison from untreated and treated persons, has identified mutations at protease positions 20, 36, 63, 71, 77 and 93 as “secondary protease mutations” in subtype B. While mutations at these positions do not cause high level drug resistance themselves, they contribute to drug resistance when present together with certain primary protease mutations, or have been shown to compensate for the decrease in catalytic efficiency caused by PI-selected primary protease mutations. The higher frequency of protease polymorphism in non-B isolates, including protease positions 20, 36, 63, 82 and 93, has raised concern that PI treatment of non-subtype B infected persons could be less effective than that of subtype B. The enzymatic efficiency of protease has a potential role in explaining drug resistance differences among subtypes. In vitro studies reporting susceptibility of clinical HIV-1 isolates of different subtypes to PIs have shown either a lack of or only small differences in susceptibility between subtypes. Although the majority of phenotypic data point to similarities in susceptibility between B and non-B subtypes, there is increasing evidence that V82I and related polymorphism in protease may play a role in subtype specific susceptibility to some PIs.

Analysis of RT region has indicates that none of the residues involved in polymerase activity are polymorphic (in the absence of antiretroviral drugs) in drug naïve persons. Comparison of non-B RT consensus sequences to the subtype B RT consensus shows differences in overall 23 residues, of which 1 (position 179) is in a position related to subtype B drug resistance. Polymorphism in non-B RT positions, which are associated with subtype B NRTI resistance include codons 69 (6% in subtype F), 75 (2% in CRF01_AE) and 118 (5, 2 and 6% in subtypes A, B and D, respectively). Mutations at position 69 (T69D/N/S/A) have been reported with exposure to all NRTIs, and contribute to NRTI resistance. Polymorphism at positions associated with subtype B non-nucleoside RT inhibitor (NNRTI) resistance is seen at codons 98 (in subtypes B, C, G), 106 (in subtypes G, CRF01_AE and CRF02_AG) and 179 (polymorphic in subtypes B, C, D, F, G, CRF02_AG, and is the consensus position in subtype A) (Table II). A98G, V106A and V179D are residues in the NNRTI binding region of the p66 subunit associated with different levels of resistance to the NNRTIs. There are accruing reports of K103N and additional major NNRTI-related drug resistance mutations in untreated persons. In addition, mutations at RT positions 98 and 179, which are considered secondary NNRTI associated mutations in subtype B, and specific adjacent polymorphisms (RT positions 177 and 178), are frequently identified among drug naive non-B infected persons.
impact of such primary and secondary mutations in RT of non-B subtypes in the developing world for clinical management of HIV-1 infection remains to be determined.

Several web-accessible sites which review antiretroviral drug resistance include the International AIDS society (IAS), USA (http://www.iasusa.org); Los Alamos Database (http://hiv-web.lanl.gov); Stanford HIV Drug Resistance Database (http://hivdb.stanford.edu); and HIV Resistance Web (http://www.hivresistanceweb.com).

**Table II.** Comparison of consensus B sequences in protease (A) and Reverse Transcriptase (B) with non-B subtypes. Columns denote subtypes and circulating recombinant forms (CRFs), and rows denote protein positions in respective protein. Sites with known drug resistance in subtype B are in Bold. Open squares are similar amino acids to consensus B sequence and shaded boxes represent polymorphism in the position compared to consensus subtype B sequence.

### (A) Protease

<table>
<thead>
<tr>
<th>Protease position</th>
<th>A</th>
<th>C</th>
<th>D</th>
<th>F</th>
<th>G</th>
<th>AE</th>
<th>AG</th>
</tr>
</thead>
<tbody>
<tr>
<td>12</td>
<td></td>
<td></td>
<td>A</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>19</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>35</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>36</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>41</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>57</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>63</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>67</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>69</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>82</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>89</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>93</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### (B) Reverse Transcriptase

<table>
<thead>
<tr>
<th>RT Positions</th>
<th>A</th>
<th>C</th>
<th>D</th>
<th>F</th>
<th>G</th>
<th>AE</th>
<th>AG</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Impact of HIV-1 subtypes on HIV transmission and disease progression**

It has been postulated that viral sequence variability may dictate biologic differences that partially explain the different epidemic patterns seen in different regions of the world. Indeed, several reports have documented that HIV-1 subtypes may differ with respect to viral load\textsuperscript{152}, chemokine co-receptor usage\textsuperscript{153}, transcriptional activation levels\textsuperscript{154} and antiretroviral drug susceptibility\textsuperscript{137-139}. While the exact functional consequences of these subtype-
related differences have not been delineated, there is some evidence that such differences may result in altered rates of disease progression\textsuperscript{155}, sexual transmission (unpublished data), and mother-to-child transmission (MTCT)\textsuperscript{156,157}.

Limited studies on subtypes and transmissibility have yielded conflicting results. A previous study in Tanzania suggested that subtypes A, C and recombinants are more likely to be perinatally transmitted than subtype D\textsuperscript{156}. A study in Uganda reported no difference in rates of perinatal transmission for subtypes A and D, whereas a large perinatal transmission study from Kenya provided a strong evidence that women infected with subtype D were more likely to transmit virus to their infants than those infected with subtype A\textsuperscript{157}. Additional studies are needed to better understand the role of subtypes in perinatal transmission.

HIV-1 subtype-specific differences in disease progression have been studied in several natural history cohorts. For instance, two studies based on incident cases have found subtype-specific differences in disease progression. A study in Senegal with limited number of patients reported that women infected with non-A subtypes were 8 times more likely to develop AIDS than those infected with subtype A\textsuperscript{155}. A large cohort study in Uganda based on more than 1000 patients demonstrated that subtype D was associated with faster progression to death and with a lower CD4 cell count during follow-up, compared with subtype A\textsuperscript{158}. However, no difference in disease progression have been found between patients infected with subtypes B and C in Israel\textsuperscript{159}, among patients infected with subtypes A, B, C and D in Sweden\textsuperscript{160}, subtypes B and CRF01_AE in Thailand\textsuperscript{161} or subtypes CRF02_AG or other subtypes in Cameroon\textsuperscript{162}. Several studies suggest faster disease progression among HIV-1 infected persons from Africa than those infected in developed countries\textsuperscript{155,163} and may presumably be due to difference in environmental factors or presence of other co-pathogens that may serve as cofactors for faster disease progression among patients in the developing world. Large cohort-based studies are needed to further confirm role of subtypes in HIV-1 transmission and disease progression.

**Implications of HIV-1 variation on vaccine design**

The importance of immune responses relevant for vaccines against the range of HIV-1 genetic variants has been the subject of intense discussion\textsuperscript{38,164-167}. HIV-1 readily mutates to escape the immune response, evolving in ways that allow it to persist in the host\textsuperscript{10,11,38,168-171}. While correlates of immune protection against HIV-1 are not clearly understood, it is known that neutralizing antibodies (nAb) correlates with protection from infection, cell-mediated immune responses correlates with protection from disease and that the effectiveness of antiviral memory T cells is dependent upon both the quality and the quantity of antigen-specific T cells\textsuperscript{11,172}. It is widely agreed that both cell-mediated (polyfunctional CD4+ and CD8+ T cell) responses and humoral responses (specifically induction of virus-neutralizing antibodies (nAbs), conferring broad and long-lasting protection against diverse HIV variants, are needed for an efficient vaccine against HIV-1\textsuperscript{11,166,172}. An ideal vaccine should include both a component to elicit CMI (polyfunctional CD4 and CD8 responses with ability to induce both gamma-interferon and interleukin-2) and one to stimulate nAbs\textsuperscript{166,167,172}. The most likely design of such a vaccine could be the use of DNA plasmid and/or a live, recombinant virus vector to evoke the former response along with a subunit protein for the latter\textsuperscript{167,173-175}.

Given the enormous worldwide diversity of HIV variants, the design and choice of antigen identities for inclusion in candidate AIDS vaccines is a daunting challenge\textsuperscript{38,174,175}. Reported preferential intraclade cellular and humoral responses and the existence of subtype-specific epitopes support the use of vaccines based on immunogens that genetically match the viruses circulating in the targeted population in which it is tested\textsuperscript{175}. Use of such vaccines is recommended by a WHO-UNAIDS Vaccine Advisory Committee and is being
implemented in several current trials\textsuperscript{166,174}. However, given the rapidity and unpredictability with which HIV-1 genetic forms may propagate in future, a vaccine protective against all major HIV-1 circulating genetic forms is desirable, which could require multivalent formulations\textsuperscript{164,175}. Indeed, emerging data suggest that HIV-1 evolutionary relationships may be more useful than regional considerations for vaccine designs\textsuperscript{164,172}. Thus, the use of consensus or ancestor sequences for a specific-subtype has been proposed in order to minimize the genetic differences between vaccine strains and contemporary isolates\textsuperscript{112,176}. This strategy could potentially enhance cross-reactivity and breadth of a vaccine response relative to any single strain. In regions where several subtypes co-circulate, inclusion of each prevalent subtype in the vaccine may improve coverage\textsuperscript{164,167,175}.

While the role of neutralizing antibodies in sterilizing immunity against HIV-1 has clearly been recognized\textsuperscript{164,172}, the induction of such response is difficult due to the envelope’s conformational complexity and sequence diversity\textsuperscript{169,172,177}. Most of the neutralization antibody sensitivity studies have shown that neutralizing serotypes do not correlate with HIV-1 genetic subtypes\textsuperscript{178}. Few studies have examined the elucidation of cross-clade neutralization by HIV-1 vaccines, since antibody responses with neutralizing capacity against primary isolates have not usually been obtained\textsuperscript{177,179-181}. While broadly neutralizing activities have been detected frequently only in long-term non-progressors, the cross-clade recognition of neutralizing activities have shown relatively restricted repertoire\textsuperscript{182}. Given the relative importance of neutralizing antibodies in protection against HIV-1\textsuperscript{166,169}, attempts are being made to devise immunogen that can potentially induce neutralizing immune responses\textsuperscript{175,177}. New findings reveal that HIV-1 protects itself from antibodies by putting up a shield of constantly shifting sugar moieties\textsuperscript{169,183}. This shield may be contributing to the poor performance of candidate HIV-1 vaccines. A recent study showing preferential heterosexual transmission of unique monophyletic viral strains that are sensitive to neutralization provides new hope for vaccine design strategies\textsuperscript{184}.

The data on the link between genetic subtypes and HIV-specific immune responses remains controversial\textsuperscript{11,165,166}. Various studies have identified both subtype-specific\textsuperscript{185-189} as well as cross-clade epitopes in HIV-1 proteins\textsuperscript{189-194}. However, intra-subtype CTL responses are usually stronger and more frequently detected than inter-subtype responses\textsuperscript{187,194,195}. This is more prominent for those responses targeted to the more variable env gene, but also to gag, pol, and nef\textsuperscript{199}. Cross-clade recognition of various CTL epitopes have led to the idea that a single subtype immunogen could elicit CTL responses protecting against other subtypes\textsuperscript{165,166,196}. This may suggest that matching HIV vaccine candidates to the prevalent HIV-1 strains might be less important for vaccines targeted at induction of T-cell responses to conserved proteins (for instance, gag and pol)\textsuperscript{164,176}.

Further, there is evidence that the genetic identities of circulating strains of HIV may continue to evolve at the population level in response to HLA-restricted immune pressures\textsuperscript{10,197}. CTL escape mutants also need to be considered for successful vaccine strategies\textsuperscript{11,170,171}. It has been proposed that mutations that ultimately result in diminished recognition of a given epitope by a host’s CD8+ T cells via loss of epitopes binding to MHC class I molecules, or interference with antigenic processing, will increase in prevalence as a result of HIV infection of individuals that share common HLA alleles\textsuperscript{10,197,199}. Thus, HIV variants may evolve in such a way that they are not recognized by the CTL responses restricted by the HLA alleles that predominate in a given human population\textsuperscript{199}. The implication of clade-specific epitopes and their role in vaccine development is presented elsewhere\textsuperscript{200}.

Are there intersubtype HIV-1 variations in host adaptation and viral fitness?

The proportions of subtypes in defined populations are not stable but are in constant flux due to new introductions of HIV-1 subtypes, changes in human behaviour, therapeutic intervention, mode of transmission, and possibly, subtype fitness\textsuperscript{12,89,90}. Over the past decade, there has been a considerable shift in the epicenter of the HIV-1 epidemic from sub-
Saharan Africa to Southern Africa, India, and Southeast Asia. Subtype C has now emerged as the predominant clade in the world due to these regional pandemics and accounts for at least half of all infections worldwide. Although subtype B likely preceded subtype C as a founder clade in India and China, most of the new infections in these countries are attributable to subtype C isolates or intersubtype B/C recombinants. Similar trends have been observed in Kenya, Tanzania, and South America (e.g., Brazil and Argentina). This rapid insurgence of subtype C may be due to a founder effect or to intersubtype fitness differences.

Any difference in ex vivo fitness likely reflects genotypic or phenotypic variations between subtypes. For example, most subtype C isolates appear to have a third NF-kB element in the long terminal repeat (LTR) as compared to the two sites found in most subtypes, which augments transcription in the presence of HIV-1 Tat protein. Direct binding of tat to NF-kB in TAR independent fashion further suggests that extra NF-kB site in subtype C may differentially regulate transcription. Moreover, a recent study indicated that subtype-specific LTR-based viral replication rates in vivo are strongly influenced by the host environment. Recently, the tat protein of HIV-1 subtype C has been shown to be defective in chemotactic activity, suggesting possible biological differences in the subtype C viruses. Studies on HIV-1 protease activity showed increased cleavage of peptide substrates by HIV-1 subtype C versus subtype B protease. This phenotypic data suggest that dominance of subtype C in the HIV-1 epidemic could be due to increased replicative capacity of subtype C isolates over other HIV-1 subtype isolates.

In addition to these phenotypic differences, subtype C appears to have preferential usage of CCR5 co-receptor and that subtype C isolates rarely switch from a NSI/R5 phenotype to an SI/CXCR4 phenotype during late disease. Analysis of co-receptor usage during disease progression has delineated that SI/X4 HIV-1 isolates dominate late in disease, whereas NSI/R5 variants are typically present during the asymptomatic phase. Further, in vivo findings suggest that NSI/R5 HIV-1 isolates out-compete the SI/X4 variants at the site of primary infection. The reasons for this are not completely understood. Recent studies have provided evidence that subtype C virus can use CXCR4, whether antiretroviral therapy is driving the emergence of CXCR4 variants remains to be determined.

The emerging data on viral fitness have indicated that subtype C may have altered fitness, thus resulting in their longer survival. Recent studies analyzing the relative fitness in a pair-wise competition indicate that subtype C isolates are likely less fit than subtype B isolates. Decreased fitness together with persistence of NSI/R5 HIV-1 isolates linked to slower disease progression could result in increased transmission time. It is evident that to survive, viruses must continue to propagate in a living host and that a low/attenuated level of pathogenesis represents a tradeoff between virulence and transmissibility. In the case of HIV-1, the continual spread of this virus is a consequence of long asymptomatic but transmissible periods following initial infection. Individuals infected with a more pathogenic (high replication) strain will progress faster to HIV-1 disease, decreasing the probability of viral transmission. In contrast, attenuated HIV-1 strains (i.e., lower replicative capacity) would in theory delay disease progression and increase the likelihood of transmission. Thus, it is likely that the poor relative fitness of subtype C isolates can fit into model for subtype C dominance in the epidemic. Whether such decreased fitness alone can account for such an explosive subtype C epidemic around the globe remains to be determined.

**Conclusion and perspective**

HIV-1 pandemic strains have diversified extensively through mutation and recombination since their initial zoonotic transmission to human beings many decades ago in central Africa. The high error rate of HIV reverse transcriptase combined with the estimated in vivo HIV-1 replication rate of ten billion new virions each day leads to extraordinary genetic diversity of HIV. Twenty seven circulating genetic forms of the HIV-1 group M are presently recognized, including 11 subtypes and sub-subtypes, and 16 circulating recombinant forms. In addition, many
unique recombinant forms (URFs) including secondary recombinant mosaic patterns also have been identified around the world. Over the past decade, there is a shift in the epicenter of the HIV-1 epidemic from sub-Saharan Africa to Southern Africa, India, Southeast Asia and China. Subtype C has now emerged as the predominant subtype in the HIV epidemic and accounts for at least half of all infections worldwide. Genotypic analyses have provided a better understanding of the molecular diversity of HIV-1, enabling the detection of emerging HIV-1 variants and improving the tracking of the epidemic worldwide.

New genetic forms, including recombinants, are being introduced in different areas of the world, changing the molecular epidemiology of the infection. Viruses resulting from recombination may be more virulent than nonrecombinant viruses due to altered fitness or cellular tropism and may have different antiretroviral drug susceptibilities. Additionally, recombinant viruses could accelerate disease progression and increase the likelihood of mother-to-child transmission and sexual transmission by increasing viral load in the blood and genital tract; however, epidemiologic evidence for such an association has not yet been established.

The rapid replication rate of HIV and its inherent genetic variation have led to the identification of many HIV variants that exhibit altered drug susceptibility. Maintaining low-to-undetectable plasma HIV RNA levels prevents progression to AIDS and minimizes the possible emergence of drug resistant HIV-1 variants. Although most clinical outcome and phenotypic reports indicate similarity between B and non-B isolates, a few studies point to potential differences. As an increasing number of individuals with non-subtype B infection access ARV therapy, these genotypic differences in non-B consensus (wild-type) sequences may result in new drug resistant forms, as well as differences in the outcomes of antiretroviral therapy. Genotypic and biochemical studies provide an explanation for potential inter-subtype differences in susceptibility; however, larger and more rigorous prospective studies are required to translate the observed inter-subtype differences in resistance patterns to more effective, long-term drug treatment strategies.

The rapid evolution of HIV within infected hosts contributes significantly to the elusiveness of this pathogen from host antiviral responses. The complex nature of HIV envelope glycoprotein that is inherently resistant to neutralization, the selective infection, progressive destruction and impaired regeneration of CD4+ T helper cells, generation of CTL escape mutants, together with high genetic diversity with continually evolving HIV variants worldwide, makes design of an effective vaccine a formidable task. The diverse HIV subtypes, CRFs, the inter-subtype mosaic genomes further exacerbate the problem of designing of relevant vaccine immunogens. Moreover, given the recent observation that HIV may exhibit a frequency of recombination diversification that is up to an order of magnitude greater than previously appreciated, the complexities of confronting HIV diversity with effective vaccine design will remain an ongoing challenge. Given the rapidity and unpredictability with which HIV-1 genetic forms may propagate in future, a vaccine protective against all major HIV-1 circulating genetic forms is desirable, which could require multivalent formulations. Indeed emerging data suggest that use of consensus or ancestor sequences for a specific subtype could potentially enhance cross-reactivity and breadth of a vaccine response relative to any single strain.

High mutation rate of HIV-1 together with selection and adaptation to most environmental changes result in the generation of viral quasispecies. HIV-1 quasispecies are evolutionary and clinically relevant since this genetic variability can respond to selective pressure (e.g., host immune system and antiretroviral therapy), including those mediated by virus specific cytotoxic T lymphocytes. Understanding the kinetics and directions of this continuing adaptation and its impact on viral fitness, immunogenicity and pathogenicity will be crucial to the successful design of effective HIV vaccines. Further, there is evidence that circulating strains of HIV may continue to evolve at the population level in response to HLA-restricted immune pressures. Such global evolution of HIV might potentially require periodic re-assessment of the composition of the best immunogens for inclusion in AIDS vaccines.
References


105. Grez M, Dietrich U, Balfe P,* et al.* Genetic analysis of human immunodeficiency virus type 1 and 2 (HIV-1 and HIV-2) mixed infections in India reveals a recent spread of


142. Erickson JW, Gulnik SV, Markowitz M. Protease inhibitors: resistance, cross-resistance, fitness and the choice of initial and salvage therapies. AIDS 1999; 13 (Suppl A) : S189-204.


162. Laurent C, Bourgeois A, Faye MA, et al. No difference in clinical progression between patients infected with the predominant human immunodeficiency virus type 1 circulating recombinant form (CRF) 02_AG strain and patients not infected with CRF02_AG, in Western and West-Central Africa: a four-year prospective multicenter study. *J Infect Dis* 2002; 186: 486-92.


Reprint requests: Dr Renu B. Lal, Division of HIV/AIDS Prevention, National Center for HIV, STD, and TB Prevention Centers for Disease Control and Prevention, 1600 Clifton Road, Atlanta, GA 30333, USA
e-mail: rbl3@cdc.gov