

Genomic diversity of human immunodeficiency viruses: The Indian scenario

PRADEEP SETH, RAJIV M. GUPTA

Department of Microbiology, All India Institute of Medical Sciences, Ansari Nagar, New Delhi

Abstract: The human immunodeficiency virus type 1 (HIV-1) strain exhibits an extraordinary degree of genetic diversity. It has been divided into 3 groups M for main, O for outlier and N for non-M non-O. The M group of viruses have been further divided into ten distinct genetic subtypes A-H and J, K and L. The diversity of HIV-1 is achieved by a high rate of mutation and replication, immune selection of resistant strains and recombination. In an infected individual, the virus exists as a population of genetically distinct, yet related genotypes called quasispecies. The distribution of HIV-1 subtypes shows a geographical predilection. Globally, subtypes A and C account for most current infections. The identification of subtypes and circulating recombinant forms provides a means of tracking the dissemination of the pandemic worldwide. The molecular epidemiological data regarding HIV infection in India is still limited. Both HIV-1 and HIV-2 are found in India. Studies revealed the presence of multiple subtypes of HIV-1, namely A, B and C, subtype C being predominant (the majority belonged to subtype C3). Even within a genetic subtype of HIV-1, the extent of genetic and antigenic diversity is enormous. However, it is observed that some features of the envelope glycoprotein structure of the virus are conserved. It would be desirable to express such conserved structures in vaccine antigens aimed at inducing a broadly reactive humoral immune response. The analysis of genetic subtypes and inter-subtype recombinant genomes is necessary to elucidate the geographical distribution and historical evolution of the strains; to detect the introduction of divergent strains; to develop subtype-specific serological techniques; to study the pattern of transmission and degree of transmissibility among variant strains; to delineate the host range, relation to disease progression and drug resistance patterns; and to design vaccines.

Introduction

More than 40 million people are infected with human immunodeficiency virus (HIV) globally. A majority of these patients experience similar symptoms regardless of the region they come from. These patients show similar responses to the same regimen of antiretroviral therapy, meaning thereby that all HIV-positive individuals are infected with an identical version of the virus. On the contrary, these are the different variants of HIV, which are members of one large family. This phenomenon is called HIV diversity. Globally circulating strains of HIV-1 exhibit an extraordinary degree of genetic diversity, which may influence various aspects of their biology such as infectivity, transmissibility and immunogenicity.

HIV-1 strains have been divided into three groups on the basis of phylogenetic analyses of complete genome sequences. The groups were originally named M for main, O for outlier, and N for non-M non-O.¹ Group M includes viruses that dominate the global epidemic,² and are responsible for more than 99% of infections.

HIV-1 genetic subtypes

The M group viruses have been divided into ten distinct genetic subtypes or clades (A through H and J, K and L).¹

Correspondence: Dr Pradeep Seth *E-mail:* pseth@aiims.aiims.ac.in

Prototype viruses representing the genetic subtypes E and I have not yet been found. The viruses originally identified as subtype E (the predominant group of viruses involved in heterosexual transmission in Thailand) and I (a small group of viruses from the Mediterranean region) are now considered inter-subtype recombinants.

Mechanisms for HIV-1 diversity

A study of isolates from the Democratic Republic of Congo indicates central Africa as the epicentre of HIV-1 diversity, with a large number of different genetic subtypes and subtype recombinants circulating. The prevalence of inter-subtype recombinant strains is increasing and creates even more HIV-1 antigenic diversity. Several recombinant viruses have now spread epidemically to establish distinct lineages.

The diversity of HIV-1 strains is achieved by: (a) the high rate of mutation, (b) high rate of virus production, (c) immune selection of resistant strains, and (d) recombination. The viral enzyme reverse transcriptase (RT), which transcribes the viral RNA into proviral DNA, is error prone. which is approximately 1 in 4000 nucleotides. In addition, HIV-1 maintains a very high replication rate producing 10 billion virus particles per day in an infected individual. Assuming that RT creates 2-point mutations during each transcription and 1 million reverse transcriptions are taking place each day, then only every single point mutation is produced daily in an infected individual. In an

infected individual, the virus exists as a population of genetically distinct, yet related genotypes called quasispecies. The diversity also greatly increases through the recombination of divergent viral strains within a population. Further, immune-mediated selection within a host also increases the HIV-1 diversity.

Recombination events between segments from two different viral strains in the same individual have resulted in the emergence of circulating recombinant forms (CRFs). Overall, CRFs constitute 10%–20% of newly characterized strains and recombination between CRFs has also been reported.³ Phylogenetic analyses in areas of high HIV-1 transmission have shown the presence of multiple mosaic genomes.⁴

Geographical distribution of HIV-1 subtypes

The distribution of HIV-1 subtypes shows a geographical predilection.⁵ Globally, subtypes A and C account for most of the current infections followed by subtype B and the inter-subtype recombinants. Subtype B is dominant in Europe, the Americas and Australia. Subtype C probably currently infects more people worldwide than any other, accounting for more than 50% of infections; it is common in southern Africa and India. Subtypes A and D infect a large number of people in central and eastern Africa. The other subtypes infect a relatively small number of people in central Africa and South America. For example, subtype F includes isolates from Brazil and Romania, and subtypes G and H include viruses from Africa, Russia and Taiwan. Subtype K, whose *env* C2-V5 sequence branched within group M but remained distinct from all known HIV-1 subtypes, was reported from the Cameroon.⁶

In recent years, the prevalence of non-B infection has markedly increased in several European countries.⁷ Although non-B infections are infrequent in North America, a study in New York identified non-B infections in a few US citizens.⁸

Even though it is the behaviour and not the occupation of an individual that determines the risk of HIV-1 infection, some groups of individuals, particularly travellers, contribute to increasing the diversity of HIV-1 worldwide. They include, in particular, immigrants, intravenous drug abusers, tourists, truck drivers, military troops and seamen. In large countries such as Russia, China, India, Brazil and South Africa, internal migrants contribute largely to the spread of HIV-1 diversity. As HIV-1 continues to spread globally, the geographical restrictions are increasingly breaking down.

Tracking the dissemination of the pandemic

As mentioned above, the identification of subtypes and CRFs provides a means of tracking dissemination of the pandemic. The world map illustrates that most of the subtypes and CRFs are present in central Africa. This distribution indirectly suggests that a few individuals initiated the earliest HIV-1 spread outside Africa.⁵ It has been documented that the initial

spread of the pandemic in a previously unaffected area is generally characterized by a founder effect, which translates into the rapid spread of a single subtype or CRF in a defined risk group (that is, subtype B in men having sex with men in North America, western Europe, and Australia, and subtype A in intravenous drugs users [IDUs] in Russia).⁹ By contrast, in Thailand, two separate epidemics occurred almost simultaneously with subtype B and CRF01 AE in two different risk groups.¹⁰⁻¹² IDUs were infected with subtype B, whereas CRF01 AE predominantly infected patients through heterosexual contact. This finding suggests that the source case for HIV-1 infection in IDUs was from Europe, the USA, or Australia, whereas the source case in the heterosexual risk group was probably derived from central Africa. Similarly, Maitra *et al.*¹³ noted in their series that infection due to subtype A in a patient with idiopathic thrombocytopenic purpura (ITP) and a husband-wife pair from Haryana was probably derived from central Africa. Little is known about the factors driving changes in the prevalence of subtypes in the same geographical areas, such as quasi absence of subtype B in Africa with the exception of homosexuals in South Africa.¹⁴ Similarly, replacement of subtype B in the northeastern region of India by subtype C is quite intriguing (Seth P, unpublished).

The Indian scenario

The scenario of HIV infection in India is extremely grim; the molecular epidemiological information is still limited. Till date, a few preliminary studies that have been conducted on these lines have all suffered from drawbacks such as small sample size, lack of proper representation from various geographical regions and inclusion of different risk groups and modes of transmission. Sahni *et al.*¹⁵ studied 125 HIV-1 seropositive individuals from different regions of India. In their study population, most of the infections were heterosexually acquired (84.8%). The other modes of transmission were blood and blood products (8.8%), IDU (3.2%), vertical transmission (1.6%) and transmission through homosexual and artificial insemination (0.8%).

Initial studies have indicated that both HIV-1 and HIV-2 are present in India.¹⁶ Subsequent studies have indicated the presence of subtype C. The subtype C isolates were found to cluster with South African isolates of NOF and ZAM 18.^{13,17} Maitra *et al.*¹³ studied the genomic diversity of HIV-1 in India by partial sequencing of the *gp120* gene (C2-V3-C3 region). The study revealed the presence of multiple subtypes of HIV-1, namely subtypes A, B and C. However, the predominant subtype was C (82% of sequences analysed). Most of the subtype C sequences obtained in India were related to the African subtype C sequences in the C2-V3-C3 region. Most of the subtype B sequences were obtained from IDUs from the northeastern state of Manipur and were related to subtype B sequences circulating in Thailand. The subtype A sequences were related to central and east African subtype sequences.

In another study,¹⁵ heteroduplex migration analysis (HMA) was used as a tool for epidemiological typing of HIV-1 isolates. A majority of the strains present in India belonged to subtype C (78.4%). Subtypes A, B and E were found in samples from the north and northeast. Further analysis of HMA data suggested that subtype C3 was the predominant strain of HIV-1 circulating in India (68%), whereas subtype C2 and C4 were present in fewer samples (8% and 2.4%, respectively). Subtype A was found in a patient with ITP who had received a blood transfusion in 1987 in Delhi and also in a husband-wife pair from Haryana where the husband had acquired infection through a female sex worker (FSW). Of the 11 patients with subtype B (Thai type subtype B), 6 acquired infection through FSW in the northeastern region of the country and 3 acquired infection through intravenous drug use in Manipur in the northeast. One husband-wife pair from Delhi was infected with subtype B' where the wife was inadvertently infected through artificial insemination in South Africa and she later infected her husband. Subtype E was found in a husband-wife pair from Uttar Pradesh in north India where the husband got infected in Manipur through a commercial sex worker (CSW) and then apparently infected his wife through unprotected intercourse. In 11 samples, HMA could not clearly identify the subtype of the envelope sequence. It is possible that these samples were recombinant HIV (homotypic or 'intra-clade' recombinants and heterotypic or 'inter-clade' recombinants). In a recent study on HIV-1 subtypes circulating in the eastern and northeastern regions of India,¹⁸ subtype C was again found to be predominant, with subtype C3 accounting for 50% and subtype C2 13%. The same study documented subtype C (68%) and B (20%) among the strains from IDUs in Manipur.

Implications of molecular heterogeneity of HIV-1

It is important to emphasize that the genetic subtypes or recombinant lineages of HIV-1 are not analogous to classic viral serotypes. The HIV-1 genetic diversity currently present in the human population dwarfs anything that has been described for other human viral infections. To put the situation into perspective: a few amino acid changes in one of the envelope glycoproteins of the influenza virus may be sufficient to trigger a new epidemic and reassortants of influenza virus envelope genes may lead to devastating pandemics.¹⁹ Yet, in HIV-1, replicating viruses can differ as much as 10% in the amino acid sequence even within a single individual. Therefore, within a genetic subtype, the extent of HIV-1 genetic and antigenic diversity is enormous when compared to the diversity found for viruses for which effective vaccines have been developed. The degree of genetic, and hence antigenic, diversity is daunting from the perspective of HIV-1 vaccine development. However, the description of a small number of human monoclonal antibodies that do neutralize many different HIV-1 isolates, including ones from different genetic subtypes, suggests that some features of the envelope glycoprotein structure are conserved.²⁰ It would, therefore, be desirable to

express such conserved structures in vaccine antigens aimed at inducing a broadly reactive humoral immune response. Knowing that genetic variants can escape immunosurveillance, information on genetic subtypes becomes important. Nonetheless, if phylogenetically defined groups broadly correspond to antigenic groups, development of vaccine using the most appropriate (antigenically related) strain will require information regarding the genetic subtypes of the regionally found strains that are most likely to pose a challenge. In addition, distribution of subtypes helps us understand the global HIV epidemic.

In Thailand, the prevalent subtypes have been shown to vary according to the risk group within the same city.¹⁰ The predominance of phylogenetically clustered strains in some geographical areas may reflect introduction by single individuals into high-risk groups.¹⁰ The prevalence of subtype and degree of inter-isolate variation in any region or group also provides us with clues about the origin and spread of HIV. Thirdly, in the face of this global variation of HIV-1, little is known about the correlation of genetic diversity to the biological properties of the virus. For example, HIV-1 subtypes B and E have caused two parallel epidemics in Thailand.¹¹ The subtype B infection, predominating in the IDU population, is different from the North American and European subtype B strains, and is generally referred to as subtype B. The subtype E strains, which are actually recombinants of subtype A and E with the *gag* gene of subtype A and the *env* gene of subtype E, are associated mostly with heterosexual transmission.¹²

The reasons for analysing genetic subtypes and inter-subtype recombinant genomes are compelling. These include: molecular epidemiological studies to elucidate the geographical distribution of strains co-circulating in a particular region; to study the historical evolution of strains in a population group; detect divergent strains being introduced; develop subtype-specific serological techniques such as synthetic peptide enzyme immunoassays for accurate diagnosis and improved sensitivity and specificity; to study the patterns of transmission and degree of transmissibility among variant strains; to delineate the host range and specificity of viral strains such as NSI and SI and their relation to disease progression; to study the drug resistance patterns among subtypes and quasispecies in an individual over time; and to design vaccines based on individual sequence subtypes. Moreover, information on the presence of multiple HIV infections caused by distinct viral subtypes circulating within a population, and the long-term observation of patients dually infected with distinct HIV strains may contribute to a better understanding of the pathogenesis of HIV infections. To study the magnitude of the HIV-1 epidemic in India, more extensive molecular epidemiological analysis of the high-risk population from different geographical regions of the country will be required to substantiate the extent of subtype representation in individual regions. Sequencing of full-length genes rather than sub-genomic fragments of the viral genome are the order of the day.

Acknowledgement

This study was supported by a research grant from the Department of Biotechnology, Ministry of Science and Technology, Government of India, New Delhi.

References

- Robertson DL, Anderson JP, Bradac JA, Carr JK, Foley B, Funkhouser RK, *et al.* HIV-1 nomenclature proposal. *Science* 2000; **288**:55–116.
- UNAIDS 2003. *Report on Global HIV/AIDS epidemic. The Barcelona report.* July 2003. [www.UNAIDS.org/barcelona/presskit/Barcelona %20report/content.html](http://www.UNAIDS.org/barcelona/presskit/Barcelona_%20report/content.html).
- McCutchan FE. Understanding the genetic diversity of HIV-1. *AIDS* 2000; **14**:S31–S44.
- Malim MH, Emerman M. HIV-1 sequence variation: Drift, shift, and attenuation. *Cell* 2001; **104**:469–72.
- Kuiken C, Thakallapalli R, Esklid A, de Ronde A. Genetic analysis reveals epidemiologic patterns in the spread of human immunodeficiency virus. *Am J Epidemiol* 2000; **152**:814–22.
- Roques P, Menu E, Narwa A, Scarlatti G, Tresoldi E, Damond F, *et al.* An unusual HIV type 1 *env* sequence embedded in a mosaic virus from Cameroon: Identification of a new *env* clade. *AIDS Res Hum Retroviruses* 1999; **15**:1585–9.
- McGarrigle CA, Nicoll A. Prevalence of HIV-1 among attenders at sexually transmitted disease clinics: Analyses according to country of birth. *Sex Transm Infect* 1998; **74**:415–20.
- Weidle PJ, Ganea CE, Irwin KL, Pieniazek D, McGowan JP, Olivo N, *et al.* Presence of human immunodeficiency virus (HIV) type 1, group M, non-B subtypes, Bronx, New York: A sentinel site for monitoring HIV genetic diversity in the United States. *J Infect Dis* 2000; **181**:470–5.
- Bobkov A, Cheingsong-Popov R, Selimova L, Ladhaya N, Kazennova E, Kravchenko A, *et al.* An HIV type 1 epidemic among injecting drug users in the former Soviet Union caused by a homogeneous subtype A strain. *AIDS Res Hum Retroviruses* 1997; **13**:1195–201.
- McCutchan FE, Hegerich PA, Brennan TP, Phanuphak P, Singharaj P, Jugsudee A, *et al.* Genetic variants of HIV-1 in Thailand. *AIDS Res Hum Retroviruses* 1992; **8**:1887–95.
- Cassol S, Weniger BG, Babu PG, Salminen MO, Zheng X, Htoon MT, *et al.* Detection of HIV type 1 envelope subtypes A, B, C and E in Asia using dried blood spots: A new surveillance tool for molecular epidemiology. *AIDS Res Hum Retroviruses* 1996; **12**:1435–41.
- Hu DJ, Dondero TJ, Rayfield MA, George JR, Schochetman G, Jaffe HW, *et al.* The emerging genetic diversity of HIV. The importance of global surveillance for diagnostic, research and prevention. *JAMA* 1996; **275**:210–26.
- Maitra A, Singh B, Banu S, Deshpande A, Robbins K, Kalish ML, *et al.* Subtypes of HIV Type 1 circulating in India: Partial envelope sequences. *AIDS Res Hum Retroviruses* 1999; **15**:941–4.
- Harmelen J van, Wood R, Lambrick M, Rybicki EP, Williamson AL, Williamson C. An association between HIV-1 subtypes and mode of transmission in Cape Town, South Africa. *AIDS* 1997; **11**:13–16.
- Sahni AK, Prasad VV, Seth P. Genomic diversity of human immunodeficiency virus type-1 in India. *Int J STD AIDS* 2002; **13**:115–18.
- Grez M, Dietrich U, Balfe P, von Briesen H, Maniar JK, Mahambre G, *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 HIV-1 and HIV-2 from a single ancestor for each of these viruses. *J Virol* 1994; **68**:2161–8.
- Voevodin A, Crandall KA, Seth P, Mufti SA. HIV type 1 subtypes B and C from new regions of India and Ethiopian expatriates in Kuwait. *AIDS Res Hum Retroviruses* 1996; **12**:641–3.
- Mandal D, Jana S, Bhattacharya SK, Chakrabarti S. HIV type 1 subtypes circulating in eastern and northeastern regions of India. *AIDS Res Hum Retroviruses* 2002; **18**:1219–27.
- Dowdle WR. Influenza A virus recycling revisited. *Bull World Health Organ* 1999; **77**:820–8.
- Trkola A, Purtscher M, Muster T, Ballaun C, Buchacher A, Sullivan N, *et al.* Human monoclonal antibody 2G12 defines a distinctive neutralization epitope on the gp120 glycoprotein of human immunodeficiency virus type I. *J Virol* 1996; **70**:1100–8.

Best Wishes From :

AVENTIS

Aventis Pharma Limited

***1113-1114, Ansal Tower, 38, Nehru Place,
New Delhi-110019, India.***