Current trends of HIV recombination worldwide

Katherine A. Lau,1 Justin J.L. Wong2
1Retroviral Genetics Division, Centre for Virus Research, Westmead Millennium Institute, Westmead Hospital, The University of Sydney; 2Gene and Stem Cell Therapy Program, Centenary Institute, Royal Prince Alfred Hospital, The University of Sydney, Sydney, Australia

Abstract

One of the major characteristics of HIV-1 is its high genetic variability and extensive heterogeneity. This characteristic is due to its molecular traits, which in turn allows it to vary, recombine, and diversify at a high frequency. As such, it generates complex molecular forms, termed recombinants, which evade the human immune system and so survive. There is no sequence constraint to the recombination pattern as it appears to occur at inter-group (between groups M and O), as well as intra- and intra-subtype within group M. Rapid emergence and active global transmission of HIV-1 recombinants, known as circulating recombinant forms (CRFs) and unique recombinant forms (URFs), requires urgent attention. To date, 55 CRFs have been reported around the world. The first CRF01_AE originated from Central Africa but spread widely in Asia. The most recent CRF; CRF55_01B is a recombinant form of CRF01_AE and subtype B, although its origin is yet to be publicly disclosed. HIV-1 recombination is an ongoing event and plays an indispensable role in HIV epidemics in different regions. Africa, Asia and South America are identified as recombination hot-spots. They are affected by continual emergence and co-circulation of newly emerging CRFs and URFs, which are now responsible for almost 20% of HIV-1 infections worldwide. Better understanding of recombinants is necessary to determine their biological and molecular attributes.

HIV-1 molecular epidemiological trends in worldwide

Since the first clinical observation of human immunodeficiency virus (HIV) in the United States in 1981, it has become a worldwide epidemic. During this spread, HIV started to diversify and undergo molecular evolution. As a result, a plethora of HIV types have been identified, which prompted further genotypic classifications. To date, HIV is divided into HIV-1 and HIV-2.1,2 HIV-1 is driving the global HIV pandemic and is further subdivided into three different groups (M, O and N). In contrast, HIV-2 is less transmissible than HIV-1 and has largely remained confined to West Africa.3 Within HIV-1 group M, nine genetically distinct subtypes are recognised (A-D, F-H, J and K), along with 55 circulating recombinant forms (CRFs), which together are accountable for more than 90% of HIV infections worldwide.4,5 The assignment of other strains of HIV as group M sub-subtypes (A1, A2, A3, F1 and F2) is still pending, since these sub-subtypes are frequently regarded as independent subtypes due to their distinct phylogenetic topology.6 The HIV-1 subtype classification takes into account all genomic regions although it was initially based on gag and/or env sequences. The intra-subtype genetic variation ranges from 15-20% whereas the variation between subtypes is approximately 25-35%, depending on the subtypes and genome regions examined.7 Intra-subtype variation continues to increase over time and evolution rates may differ according to subtype.8,9 Although the global availability of sequencing techniques has increased, the classification of HIV-1 strains into subtypes and CRFs remains a complex issue. Definitions of different HIV-1 subtypes are subject to change as new viral sequences are submitted and/or env sequences. Despite the complexity of HIV-1 subtype classification, different HIV-1 subtypes have distinct global distribution patterns. Prior to 2005, HIV-1 subtype C or recombinant forms containing at least the env gene of subtype C were responsible for 50% or more of all HIV-1 infections worldwide.10,11 This was partly a result of the pandemics in Southern Africa, South America and Asia.9,12,13 However, in recent years, HIV-1 subtype B has become the most predominant strain worldwide, with 60.1% infections (based on the subtype distribution available from the Los Alamos HIV Database; http://www.hiv.lanl.gov/). Other main subtypes such as C, A, D, G, and F have accounted for 13.8, 6.9, 3.7, 1.1, and 1.0% of HIV-1 infections, respectively. In terms of worldwide distribution of the recombinant forms, three most important CRFs (CRF01_AE, CRF_02_AG and CRF07_BC) have contributed towards 4.6, 2.9, and 0.9% of total HIV-1 infections (Figure 1A).

According to the Joint United Nations Programme on HIV and AIDS (UNAIDS), 34 million people was estimated to be living with HIV around the world in 2011. In the same year, there were an estimated 2.5 million new HIV infections. In comparison to 1999, there was a decline of 19% in the number of newly infected individuals with HIV. Sub-Saharan Africa remains the most affected area of the world; being home to 67.6% of all individuals living with HIV and accounted for 72% of Acquired Immunodeficiency Syndrome (AIDS) deaths in 2009. Within the sub-region, South Africa remains the country with the world largest HIV epidemic with 5.6 million of HIV infections and 17.2% of the total AIDS mortality in 2009. In other regions such as Asia, the number of people living with HIV remains stable and is estimated to be around 4.9 million. South and Southeast Asia are predicted to have the highest HIV infection prevalence (estimated to be at 4.1 million in 2009), due to their disparate epidemic trends. Other regions, for instance North America, Central and South America, as well as Eastern Europe and Central Asia, all accounted for 1.5, 1.4 and 1.4 million HIV infections, respectively. Overall, the global growth of AIDS epidemic appears to have stabilised. This trend is attributed to a combination of factors, including the significant scale-up of antiretroviral therapy over the past few years, the impact of HIV prevention efforts and the natural course of HIV epidemics.

HIV recombination

HIV possesses high recombination rates, due to the highly recombinogenic reverse transcriptase (RT) enzyme.12 RT has a high error rate and low binding affinity,13 which are necessary for strand transfers of reverse transcrip-
As a result, it contributes towards one of the major characteristics of HIV-1 to enormously vary genetically to form an extensive heterogeneity. Ability to recombine, together with high viral replication and genomic mutation (insertions and deletions) rates have resulted in viral population with radical genetic diversity within an individual, also known as dual-infection. This is especially common in HIV-1 infections.17,18

Recombination is not limited by sequence similarity. It has been demonstrated to occur inter-group (between HIV-1 group M and O),19 and inter- and intra-strain within HIV group M.20 These new recombinants display discrete breakpoints that can be identified between genomic regions with different phylogenetic associations.21,22 Given that the physical sites of recombination are distributed randomly along the viral genome, preferred sites for HIV-1 recombination remain unknown. Recombination becomes a common occurrence among different HIV-1 strains, and the inter-subtype recombination is the most frequently observed, although intra-subtype recombination is possible.

At the molecular level, HIV-1 recombination occurs at two different stages; during the synthesis of the minus strand DNA or plus strand DNA using different mechanisms. It is a result of the placement of multiple selection factors, which are summarised into three major processes. The first factor includes the reverse transcription step that generates recombinant genomes, dependent on the mechanism of copy choice. The second factor is the selection for the functional forms, while the last factor is the selection for the replicable forms within the host, which subsequently transmits from one host to another. All these factors are fundamental in shaping the HIV-1 predominance in the population. Thus, it becomes vital to gain further understanding and knowledge into the virus-host interplay.

HIV recombinants in the epidemic

The highly unequal geographic distribution of viral variants is the result of global variation in the HIV-1 strains, the dynamic nature of the HIV-1 epidemic, and accidental epidemiologic transmissions. The recombinant HIV-1 strains are emerging at a high frequency due to co-circulation of multiple HIV-1 subtypes in almost all geographic regions of the globe and co- or super-infection of individuals with multiple subtypes. These factors result in HIV CRFs, defined as characteristic full-length or near full-length HIV sequence that are found in three epidemiologically unlinked individuals, or URFs, when these criteria are not met. Novel and newly identified CRFs are named in the order in which they are reported and described, and thus do not represent their historical evolutionary or chronological order. When there are three or more different HIV strains forming a CRF, it is given the extension cpx for complex. CRFs that consist of five or more HIV-1 subtypes, for instance CRF18_cpx, CRF27_cpx, CRF37_cpx, and CRF49_cpx can be true mosaics, as their sequence segments are composed of many different HIV-1 subtypes or other unidentified sequences.

In many regions around the world, there is a significant displacement of the existing HIV-1 subtype by other new strains, for instance subtype B and CRF01_AE in Southern China,12,13 Southeast Asia,21,22 and subtype B and F in Brazil.24,33 It is likely that the initial predominance of a certain HIV subtype is attributable to its specific mode of transmission within a given geographical region, for instance displacement of subtype C by subtype B.32 However, there is no substantial evidence to suggest that HIV-1 strains such as CRF01_AE and subtypes A, B, C and D are more or less transmissible by a specific route, in a specific ethnic group, or in a specific cell type.35-39 As the prevalence of variants of HIV-1 recombinant viruses, either CRFs or URFs are becoming more frequent in certain regions around the world, subtype displacement is no longer due to the emergence of another pure HIV-1 subtype. HIV-1 recombinants are most prevalent in areas where multiple subtypes co-circulate and therefore play a major role in the global AIDS epidemic. To date, 55 CRFs have been identified based on the Los Alamos HIV Database, and the global pandemic of HIV-1 is likely to be affected by widespread of the

Figure 1. Worldwide distribution of HIV-1 recombinant forms. (A) Frequency of recombinant forms that contributed towards total HIV infections as of the year 2012. (B) Frequency of different recombinant forms that contributed towards 55 CRFs identified to date.
recombinant forms. Their incidence in global epidemics has been estimated around 18–20%, as reported by Hemelaar et al. and Buonaguro et al. in 2007. There is a gradual replacement and phasing-out of the initial predominant HIV-1 strains (the pure subtypes) by the increasingly-epidemiologically important CRFs, as evident from the epidemics in some countries or regions, for instance Malaysia, Brazil, and West Africa.

Identifications of the initial CRFs, namely CRF01_AE and CRF02_AG were controversial due to the constraint of HIV sequence availability and the difficulty in determining recombination breakpoints. CRF01_AE is defined as a HIV recombinant with insertion of subtype E segment at the ene region within the backbone of subtype A. Some have suggested that CRF01_AE is in fact subtype E, due to the fact that pure subtype E lineage has never been found. However, it is possible that CRF01_AE is a result of an early event of HIV-1 recombination, which is certainly plausible owing to frequent incidence of unique and unclassifiable viral sequences found in West and Central Africa. Likewise, CRF02_AG was proposed to be a pure subtype, although it was later confirmed as a CRF composed of subtype A and G.

CRF01_AE and CRF02_AG were found to form and emerge at the early stage of HIV-1 epidemic in Central Africa. Subsequently, they spread and expanded to other regions and thus play a crucial role in bringing changes in the regional and global HIV-1 epidemic trend. Today, CRF01_AE has become the most predominant CRF (Figure 1A). It is circulating mostly in Southeast Asia and is responsible for almost 5% of the total HIV-1 infection worldwide. It has also recombined with many other HIV-1 strains and contributes to a total of 9% of all currently identified CRFs, as documented in the Los Alamos HIV Database. The most prevalent pure HIV-1 subtype known to recombine with CRF01_AE is subtype B, which resulted in CRF15_01B, CRF33_01B, CRF34_01B, CRF40_01B, CRF51_01B, CRF32_01B, CRF33_01B, CRF54_01B, and CRF55_01B. There is also an inter-CRF recombination between CRF01_AE and CRF02_AG, as evident in CRF36_cpx and CRF37_cpx. Both of them are complex CRFs and consist of at least four HIV-1 strains. At present, CRF02_AG accounts for 2.9% of total global HIV-1 infections and has a relevant epidemiological prevalence. It is not confined within a geographical location and has been reported to co-circulate with many other HIV-1 strains within the same region. However, CRF02_AG is only found in 3% of all 55 CRFs that have been reported in the Los Alamos HIV Database. This could be explained by its better capacity to prevent its virion from recombining with another HIV-1 strains, in comparison to CRF01_AE. The factor(s) underlying the ability of HIV-1 strains to recombine is an interesting topic to explore further, in order to understand molecular and biological factors that are involved within certain HIV-1 strains to undergo inter-strain recombination.

Based on the Los Alamos HIV Database, HIV-1 subtype B is the strain that has the highest recombination frequency (19%) to form novel CRF, in comparison to other HIV-1 group M viruses (Figure 1B). This is strongly indicated by its predominance worldwide. Surprisingly, subtype C, which accounts for 13.8% of total global HIV-1 infections, only have a recombination occurrence of 4% of the total 55 CRFs. On the other hand, subtype G which only accounts for 1.1% HIV-1 infection worldwide contributes up to 12% of total CRFs identified so far. The ability of subtype G to combine could be due to the formation of heterozygous virions or functional chimeras with another HIV-1 strain to form a new and novel CRF. This also provides evidence for the possible generation of new viruses that possess biological properties distinct to parental strains, which may in turn lead to new epidemiological trend.

Distribution of HIV recombinant forms

HIV-1 recombination usually takes place in regions where different subtypes and CRFs cocirculate. These regions are named recombi-
nation hot-spots. Recombination is an ongoing process while HIV-1 recombinants continue to evolve, emerge and disseminate into world population in an aggressive manner, to give rise to new CRF or URF.

New and novel URFs are highly prevalent, and are markedly evident in Africa, South America, Cuba, China and Southeast Asia. They can continue to spread in the population, and lead to the emergence of new CRFs. This phenomenon has been demonstrated by the emergence of several CRFs that are derived from unique HIV-1 subtype B and F-like URFs in South America. The most recent occurrence is evident in the discovery of an URF strain 07MYKLD49, which was first reported in 2007 by our group, and later classified with two other URFs as a new CRF in 2012. More second generation recombinants have emerged due to the recombination of CRFs with other HIV-1 strains, indicating the up-coming of new recombinant strains. Their emergence has contributed towards a significant epidemiological impact, as well as a continual, active transmission of URFs within the region. It has been 20 years since the first CRF was reported and the high frequency of rapidly emerging CRFs complicates the current HIV-1 epidemic trend. Based on the information available for all CRFs sequences submitted to the Los Alamos HIV Database, we notice that new CRFs continue to emerge and increase steadily worldwide, especially from 1999 onwards (Figure 2A). Three geographical distinct regions; Africa, Asia and South America have been identified as regions with the highest CRF prevalence. Apart from the increasing number of new CRFs, Africa, in particular Central Africa, has incredibly high proportions of URFs and other unclassifiable HIV sequences. This same region is also the home and origin to all HIV-1 group M pure subtypes, which were found at the early stage of the HIV-1 epidemic. For the first five years since the first report of CRFs, Africa dominates the entire worldwide HIV-1 CRF prevalence, followed by Asia (Figure 2B). This phenomenon is bias towards African and Asian regions, as there is no other novel CRF found elsewhere. In the following five years between the period of 1998 and 2002, South America, where no prior evident of report of CRF, has recorded a significant cumulative percentage of new CRFs, for the first time. Africa remains the region that leads the HIV-1 CRF epidemic trend by demonstrating a comparable cumulative percentage of new CRFs reported to that found elsewhere (excluding Asia and South America) during that period (Figure 2B). However, there was a turning point in the world HIV-1 CRF epidemic trend that took place in the subsequent 10 years (between 2003 and 2012). Although there is a marked increase in the percentage of newly discovered CRFs observed in the last 10 years, Asia and South America replaced Africa as regions with the highest cumulative percentages of CRF reported; 70% and 80% respectively (Figure 2B). This is mainly due to the increase in the co-circulation of multiple HIV-1 strains within these regions. It is also noteworthy that during this explosive period of the global HIV-1 CRF expansion (between 2003 and 2012), less novel CRFs have emerged from Africa, particularly Central Africa. This observation suggests a stable and more mature epidemic, consistent with it being the centre of the HIV-1 group M epidemic.

The current trend in the distribution of global HIV-1 strains reflects its relation to multiple factors, including founder effects, active urbanisation, expanding population, escalating transport network, high migration rate, high-risk behaviours, as well as cultural and sexual-related issues. It has also been predicted that the active and continuous emergence of novel CRFs is linked to the biological advantages of certain parental strains, although the extent of these advantages in their differential spread is still uncertain.

Conclusions

The divergence of HIV-1 strains, resulting in the emerging of HIV-1 recombinant forms has given rise to a range of impact in every aspects of the HIV pandemic. New HIV-1 strains have emerged due to either neutral or deleterious recombination of HIV-1 genome from different parental subtypes and thus, are very unlikely to confer beneficial properties to the HIV-1 pandemic worldwide. It can be envisaged that the recombination pattern will continually be driven by the rapid removal of mutations that give negative impacts to viral replication. Concurrently, it will also bring forward the strain that acquire the highest adaptive capacity in a given environment. In conclusion, we postulate that apart from the pure HIV-1 subtypes (B, C and A), the most prevalent CRFs for instance, CRF01_AE may be responsible in shaping the global HIV-1 distribution trend. The current ongoing phylogenetic analysis of HIV-1 sequence variability, which further led to the study of the origin, evolution and spread of HIV-1 recombinants have become more extensive and active in the past five to ten years. It is encouraging that the impact of HIV-1 recombination on diagnosis, antiretroviral treatment, transmission, pathogenesis, clinical management and vaccine development is widely recognised. Ultimately, a closer look into approaches to understand the central role of HIV-1 recombination in the HIV-1 pandemic is essential to provide better insights into the development of next generation HIV vaccines.

References

1. Coffin J, Haase A, Levy JA, et al. What to call the AIDS virus? Nature 1986;321:10.
2. Reeves JD, Doms RW. Human immunodeficiency virus type 2. J Gen Virol 2002; 83:1253-65.
3. Clavel F. HIV-2, the West African AIDS virus. AIDS 1987;1:135-40.
4. McCutchan FE. Understanding the genetic diversity of HIV-1. AIDS 2000;14 Suppl 3:S31-44.
5. Peeters M, Sharp PM. Genetic diversity of HIV-1: the moving target. AIDS 2000;14 Suppl 3:S129-40.
6. Robertson DL, Anderson JP, Bradac JA, et al. HIV-1 Nomenclature proposal: a reference guide to HIV-1 classification. In: Kuiken CLFB, Hahn B, Marx PA, et al, eds Los Alamos, NM: Theoretical Biology and Biophysics Group, Los Alamos National Laboratory; 1999.
7. Korber B, Gaschen B, Yusim K, et al. Evolutionary and immunological implications of contemporary HIV-1 variation. Br Med Bull 2001;58:19-42.
8. Rambaut A, Robertson DL, Pybus OG, et al. Human immunodeficiency virus. Phylogeny and the origin of HIV-1. Nature 2001; 410:1047-8.
9. Abecasis BA, Vandamme AM, Lemey P. Quantifying differences in the tempo of human immunodeficiency virus type 1 subtype evolution. J Virol 2009;83:12917-24.
10. Essex M. Human immunodeficiency viruses in the developing world. Adv Virus Res 1999;53:71-88.
11. Hemelaar J, Gouws E, Ghys PD, Osmanov S. Global and regional distribution of HIV-1 genetic subtypes and recombinants in 2004. AIDS 2006;20:W13-23.
12. Luo CC, Tian C, Hu DJ, et al. HIV-1 subtype C in China. Lancet 1995;345:1651-2.
13. Piyaraksilp S, McCutchan FE, Carr JK, et al. A recent outbreak of human immunodeficiency virus type 1 infection in southern China was initiated by two highly homogeneous, geographically separated strains, circulating recombinant form AE and a novel BC recombinant. J Virol 2000;74:11286-95.
14. Soares EA, Martinez AM, Souza TM, et al. HIV-1 subtype C dissemination in southern Brazil. AIDS 2005;19 Suppl 4:881-6.
15. Hu WS, Temmink HM. Retroviral recombination and reverse transcription. Science 1999;2501227-33.
16. Preston BD, Poiesz BJ, Loeb LA. Fidelity of HIV-1 reverse transcriptase. Science 1988;242:1168-71.
17. Levy DN, Aldrovandi GM, Kutsch O, Shaw GM. From the cover: dynamics of HIV-1
Review

Fernandez JC, Morgado MG. Identification of two new CRF in Rio de Janeiro State, Brazil. AIDS 2008;22:433-5.

31. Sanabani SS, Pastena ER, Neto WK, et al. Characterization and frequency of a newly identified HIV-1 BFI intersubtype circulating recombinant form in Sao Paulo, Brazil. Virol J 2010;7:74.

32. Walker PR, Pybus OG, Rambaut A, Holmes EC. Comparative population dynamics of HIV-1 subtypes B and C: subtype-specific differences in patterns of epidemic growth. Infect Genet Evol 2005;5:199-208.

33. Pope M, Frankel SS, Mascola JR, et al. Human immunodeficiency virus type 1 strains of subtypes B and E replicate in cutaneous dendritic cell-T-cell mixtures without displaying subtype-specific tropism. J Virol 1997;71:8001-7.

34. Dittmar MT, Simmons G, Hibbitts S, et al. Langerhans cell tropism of human immunodeficiency virus type 1 subtype A through F isolates derived from different transmission groups. J Virol 1997;71:8008-13.

35. Renjifo B, Fawzi W, Mwakagle D, et al. Differences in perinatal transmission among human immunodeficiency virus type 1 genotypes. J Hum Virol 2001;4:16-25.

36. Ball SC, Abrah A, Collins KR, et al. Comparing the ex vivo fitness of CCR5-tropic human immunodeficiency virus type 1 isolates of subtypes B and C. J Virol 2003;77:1021-38.

37. Yang R, Xia X, Kusagawa S, et al. Preferential in-utero transmission of HIV-1 subtype C compared to HIV-1 subtype A or D. AIDS 2004;18:1629-36.

38. Eschleman SH, Guay LA, Mwatha A, et al. Comparison of mother-to-child transmission rates in Ugandan women with subtype A versus D HIV-1 who received single-dose nevirapine prophylaxis: HIV Network for Prevention Trials 012. J Acquir Immune Defic Syndr 2005;39:593-7.

39. Hemelaar J, Gouws E, Gheys PD, Osmanov S. Global trends in molecular epidemiology of HIV-1 during 2000-2007. AIDS. 2011;25:679-89.

40. Buonaguro L, Tonnesello ML, Buonaguro FM. Human immunodeficiency virus type 1 subtype distribution in the worldwide epidemic: pathogenetic and therapeutic implications. J Virol. 2007;81:10209-19.

41. Anderson JP, Rodrigo AG, Learn GH, et al. Testing the hypothesis of a recombinant origin of human immunodeficiency virus type 1 subtype E. J Virol 2000;74:10752-65.

42. Vidal N, Peeters M, Mulanga-Kabeya C, et al. Unprecedented degree of human immunodeficiency virus type 1 (HIV-1) group M genetic diversity in the Democratic Republic of Congo suggests that the HIV-1 pandemic originated in Central Africa. J Virol 2000;74:10498-507.

43. Abecasis AB, Leme P, Vidal N, et al. Recombination confounds the early evolutionary history of human immunodeficiency virus type 1: subtype G is a circulating recombinant form. J Virol 2007;81:8543-51.

44. Carr JK, Avila M, Gomez Carrillo M, et al. Unprecedented degree of human immunodeficiency virus type 1: subtype G is a circulating recombinant form. J Virol 2007;81:8543-51.
South America. AIDS 2001;15:F41-7.

55. Ng KT, Ong LY, Takebe Y, et al. Genome sequence of a novel HIV-1 circulating recombinant form 54_01B from Malaysia. J Virol 2012;86:11405-6.

56. Kalish ML, Robbins KE, Pieniazek D, et al. Recombinant viruses and early global HIV-1 epidemic. Emerg Infect Dis 2004;10:1227-34.

57. Worobey M, Gemmel M, Teuwen DE, et al. Direct evidence of extensive diversity of HIV-1 in Kinshasa by 1960. Nature 2008;453:661-4.

58. Gray RR, Tatem AJ, Lamers S, et al. Spatial phylodynamics of HIV-1 epidemic emergence in east Africa. AIDS 2009;23:F9-F17.

59. Tee KK, Saw TL, Pon CK, et al. The evolving molecular epidemiology of HIV type 1 among injecting drug users (IDUs) in Malaysia. AIDS Res Hum Retroviruses 2005;21:1046-50.

60. Lau KA, Wang B, Miranda-Saksena M, et al. Evidence for possible biological advantages of the newly emerging HIV-1 circulating recombinant form from Malaysia - CRF33_01B in comparison to its progenitors - CRF01_AE and subtype B. Curr HIV Res 2010;8:259-71.

61. Tee KK, Kusagawa S, Li XJ, et al. Isolation and characterization of a replication-competent molecular clone of an HIV-1 circulating recombinant form (CRF33_01B). PLoS One 2009;4:e6666.