Yutaka Takebe, Rie Uenishi, and Xiaojie Li

Laboratory of Molecular Virology and Epidemiology, AIDS Research Center, National Institute of Infectious Diseases, Tokyo 162-8640, Tokyo, Japan

Global Molecular Epidemiology of HIV: Understanding the Genesis of AIDS Pandemic

I. Chapter Overview

Global dissemination of the *Human immunodeficiency virus* (HIV) represents a dramatic and deadly example of recent genome emergence and expansion. Since HIV-1 group M began its expansion in human population roughly 70 years ago (in early twentieth century), it has been diversifying rapidly, now comprising a number of different subtypes and

circulating recombinant forms (CRFs). Molecular epidemiological method has been useful tool to analyze the origin of HIVs and to track a course of global HIV dissemination. It could also provide the information critical to prevention and future vaccine strategies. In this chapter, we describe the classification and distribution of HIV genotypes and the biological and public health implications of genetic variability of this deadly pathogen.

II. Introduction

The HIV/AIDS pandemic continues to expand globally at a rate of 13,000 new infections everyday. The Joint United Nations Program on HIV/AIDS (UNAIDS) estimates that 40.3 (36.7–45.3) million individuals are living with HIV/AIDS, and about 25 million patients have already died (UNAIDS/WHO, 2005). A total of estimated 65 million individuals have been thus infected with HIV worldwide since the epidemic started a quarter century ago. In 2005 alone, there were 4.9 (4.3–6.6) million new HIV infections and 3.1 (2.8–3.6) million AIDS deaths (UNAIDS/WHO, 2005). This could be translated as that 9.3 new infections and 5.9 AIDS deaths occurred every minute (or a new infection every 6–7 sec and an AIDS death every 10 sec) worldwide. Figure 1 illustrates the magnitude of HIV/AIDS epidemic in different regions of the world.

Heterosexual transmission remains the dominant mode of transmission and accounts for ~85% of all HIV infections worldwide. Sub-Saharan Africa is an epicenter of the pandemic and continues to have high rates of new infections [3.2 (2.8–3.9) million per year]. It accounts for ~65% of new infections occurred worldwide in 2005 (Fig. 1). While HIV/AIDS epidemics came later in Asia, Asia is becoming the epicenter of second largest epidemic with ~1 million infections annually, accounting for 20% of new infections in the world (Fig. 1). Outside of Sub-Saharan Africa, one third of HIV infections are acquired through injecting drug use, most of which (an estimated 8.8 millions) are in Eastern Europe and central and Southeast Asia. The interplay between injecting drug use and unprotected sex fuels the epidemics in many countries in Asia (Fig. 1).

Molecular epidemiology has been a useful tool to analyze the origin of HIVs and to track a course of global HIV spread. The study areas include the distribution of HIV genotypes in different geographic areas, route of global and regional virus spread, molecular features of emerging epidemics and regional outbreaks, and specific association with different epidemiologic features, such as risk behaviors. Recent investigations also provide the new data on the role of recombination in the generation of HIV genetic diversity and the frequency of dual and superinfections. In this chapter, we overview the recent advances in the study of global molecular epidemiology of HIV and discuss its biological and public health implications.

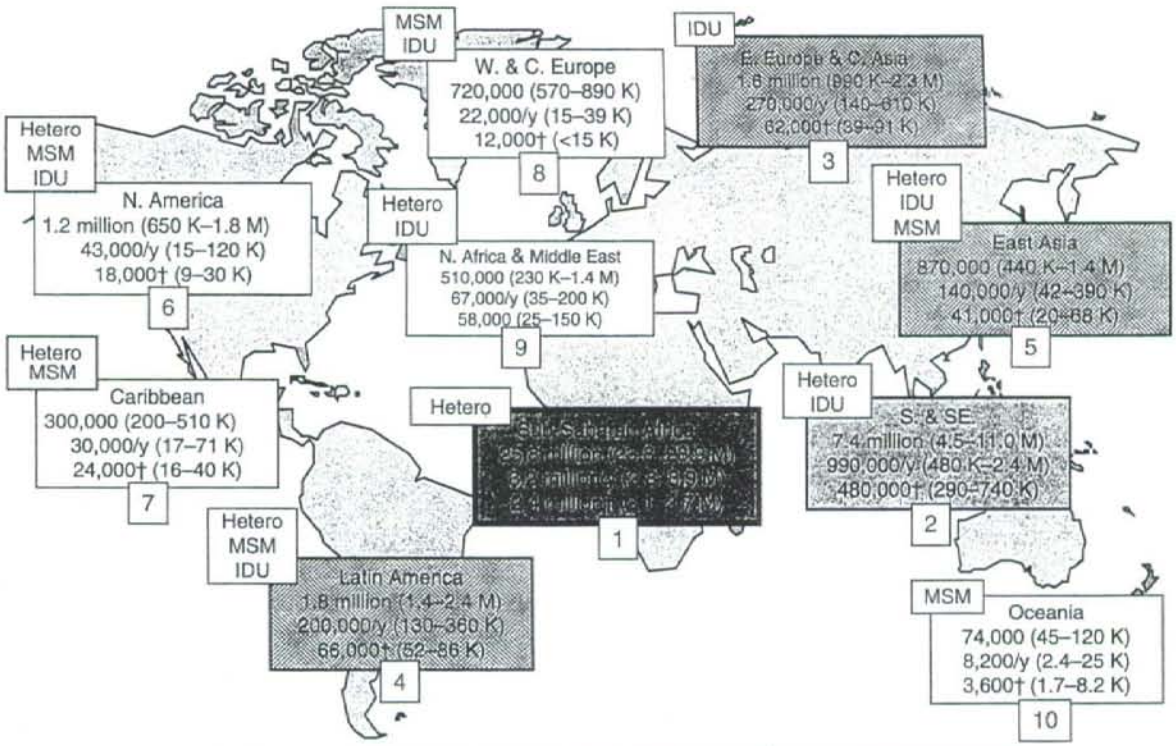

FIGURE 1 Distribution and estimated magnitude of HIV infections in different geographical regions of the world. UNAIDS/WHO estimates for the number of people living with HIV, people newly infected with HIV (/y) and AIDS death (†) in 2005 for the respective geographical regions are boxed. Predominant modes of transmission for each region are shown in the left upper box: Hetero, heterosexual; MSM, men who have sex with men; IDU, injecting drug user. Arabic numeral "1" in the box indicates the most afflicted region with the highest annual incidence and "10" indicates the least afflicted region. Global totals are shown at the bottom. Illustrations based on UNAIDS/WHO (2005).

III. Genotype Classification of HIVs

A. HIV Types (HIV-1 and HIV-2)

Etiologic agents for AIDS are subdivided into two related human retroviruses (lentiviruses): HIV-1 (HIV type 1) and HIV-2 (HIV type 2). HIV-1 is distributed worldwide, accounting for the majority of HIV infections. By contrast, HIV-2 is confined to West Africa and southern/western India (Schim van der Loeff *et al.*, 1999). Sporadic occurrences and transmission outbreaks of HIV-2 have been reported from many European countries (Cilla *et al.*, 2001; Damond *et al.*, 2004) and North and South America (Sullivan *et al.*, 1998) as well as Korea (Kim *et al.*, 2000; Nam *et al.*, 2006). It is known that the sexual and perinatal transmissions of HIV-2 are much less efficient than HIV-1 (Kanki *et al.*, 1994). This is attributed to a lower viral burden of HIV-2 during the relatively long asymptomatic period and may be the reason why the number of HIV-2-infected individuals has remained small, confined to limited geographic regions compared with HIV-1 infections (Schim van der Loeff *et al.*, 1999).

B. Genotype Classification of HIV-1s

Phylogenetic sequence analyses of HIV-1 strains distributed worldwide have identified three distinct groups of HIV-1 (M, N, and O), and nine genetic subtypes (A–D, F–H, J, and K) within major group (M) (Robertson *et al.*, 1999) (Fig. 2). The vast majority (more than 95%) of HIV-1 strains belong to group M (for Major or Main). Group O (for Outlier) comprises a pool of highly divergent, genetically related strains (Charneau *et al.*, 1994; Gurtler *et al.*, 1994; Loussert-Ajaka *et al.*, 1995; Vanden Haesevelde *et al.*, 1994) (Fig. 2). Group O infections are limited to people living in central Africa (Cameroon, Gabon, and equatorial Guinea), but even in this area they represent a small minority of HIV-1 infections. Only a few cases of group N (for New, or non-M/non-O) infections were identified in only limited number of patients from Cameroon (Simon *et al.*, 1998). HIV-1 group N infections fail to react serologically in standard whole-virus enzyme-linked immunosorbent assay (ELISA), yet are readily detectable by conventional Western blot analysis.

1. HIV-1 Subtypes

HIV-1 group M viruses are classified into at least nine discrete genetic subtypes (A–D, F–H, J, and K) based on the sequence of complete viral genomes (Fig. 2). In some subtypes, subclusters within subtypes were identified, leading to a classification into subsubtypes: The subtype F is subdivided into two subsubtypes: F1 and F2, and subsubtypes A1 and A2 strains were identified within subtype A (Fig. 2). The subtypes B and D are more

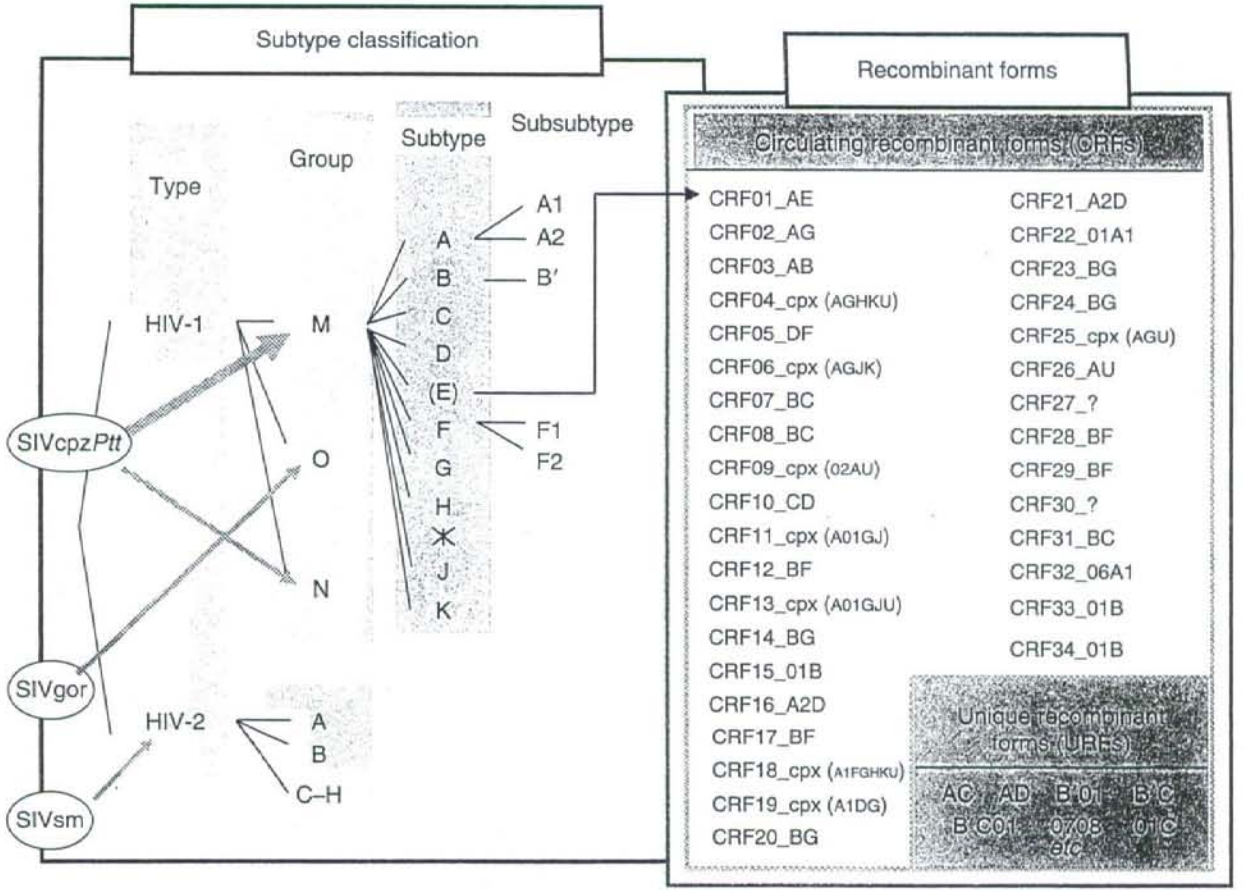

FIGURE 2 Classification of HIV genotypes and their origins. HIVs are classified based on the following four different strata: types (types 1 and 2), groups (M, O, and N for HIV-1; A–G for HIV-2), subtypes (A–K for HIV-1), and subsubtypes (A1 and A2, F1 and F2). HIV-1 recombinants in pandemic strains belonged to group M are categorized into circulating recombinant forms (CRFs) and unique recombinant forms (URFs) by their magnitude of dissemination. HIV-1 subtype B′ (B "prime") (Thailand variant of subtype B, also referred to as Thai-B), a unique regional subtype variant strongly associated with bloodborne transmission (prevalent among IDUs and former plasma donors/paid blood donors) in Southeast and East Asia. The plausible origins and routes of cross-species transmissions of HIVs are depicted in the left side (see text).

closely related to each other than to other subtypes, and therefore better considered as subsubtypes within a single subtype, rather than different subtypes. However, for the consistency with earlier published works, their original designation as subtypes is retained (Robertson *et al.*, 1999). Intrapatient genetic diversity of HIV-1 can vary from 6 to 10% in nucleotide sequence. HIV-1 isolates within a subtype may exhibit nucleotide distances of 15% in *gag* and up to 30% in *env* gp120-coding sequences. Intersubtype genetic diversity may range between 30 and 40%, depending on the gene analyzed. Similarly, the amino acid distances among different subtypes of HIV-1 group M reach ~25–30% in the *env* gene sequence and 15% in the *gag* gene sequence (Robertson *et al.*, 1999).

2. *HIV-1 Recombinants*

a. Circulating Recombinant Form It was realized that certain HIV-1 strains clustered with different subtypes in different regions of their genomes. Some of these mosaic HIV-1 genomes have been identified in several, apparently unlinked, individuals and play a major role in the global AIDS pandemic, designated as "CRFs" (Carr *et al.*, 1998). A total of 34 CRFs are currently recognized (http://hiv-web.lanl.gov/CRFs/CRFs.html) (Fig. 2). The global distribution of HIV genotypes are shown in Figs. 3 and 4. Under new nomenclature proposals, each CRF is designated by an identifying number, with letters indicating the subtypes involved. If the genome contains sequences originating more than two subtypes, the letters are replaced by "cpx," denoting "complex." To define a new subtype, subsubtype, or CRF, the representative strains must be identified in at least three epidemiologically unlinked individuals. Three near full-length genomic sequences are preferred, but two complete genome with partial sequences of a third strain are sufficient to designate a new subtype, subsubtype, or CRF (Robertson *et al.*, 1999).

b. Unique Recombinant Form In addition to CRFs, various types of "unique" recombinant forms (URFs) have been reported, currently without evidence of epidemic spread (McCutchan, 2000). URFs are diverse forms of HIV-1 intergenotype recombinants with unique mosaic structures, seen only in a single person or in a few epidemiologically linked individuals. Most of the URFs have been detected in the regions where multiple subtypes are cocirculating. A wide variety of URFs have been reported in the regions including Democratic Republic of Congo (DRC) (A/G/J and F1/K/U) (Vidal *et al.*, 2000), Tanzania (A1/C and A1/D) (Hoelscher *et al.*, 2001), Argentina (B/F) (Thomson *et al.*, 2000), Cuba (various combinations between subtypes A, B, D, G, and H) (Cuevas *et al.*, 2002), Spain (B/G) (Thomson *et al.*, 2001), India (A/C) (Lole *et al.*, 1999), Thailand (CRF01_AE/B) (McCutchan, 2000), Myanmar (various combinations between subtypes B′, C and CRF01_AE) (Motomura *et al.*, 2000, 2003; Takebe *et al.*, 2003), and China (B′/C) (Yang *et al.*, 2002, 2003 (Fig. 3). The detection of substantial

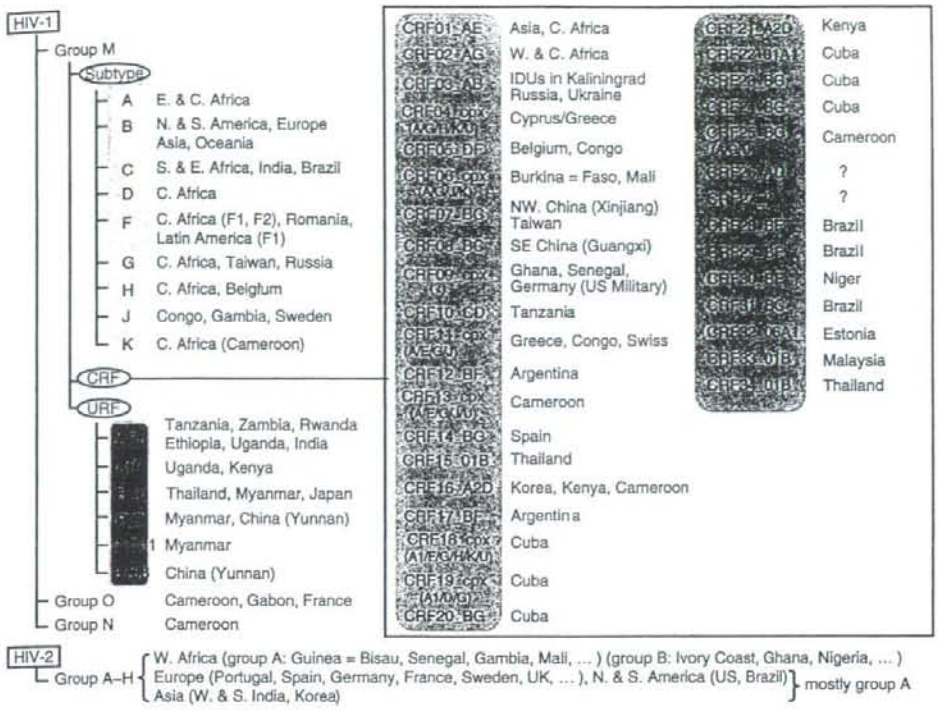

FIGURE 3 Geographical distribution of HIV genotypes. The global distributions of HIV-1 genotypes (groups, subtypes/subsubtypes, and CRFs/URFs) and HIV-2 groups are shown.

numbers of different URFs worldwide suggests that dual/multiple infections or superinfections with different lineages of HIV-1 strains might not be a rare event.

C. HIV-2: Genotype Classification and Geographic Distribution

HIV-2s are known to be members of a broader HIV-2/*Simian immunodeficiency virus* sooty mangabey (SIV-sm) phylogenetic group. HIV/SIV-sm phylogenetic groups are classified into eight genetic "groups" (A–H) (Damond *et al.*, 2004). Because these HIV-2 clades are nearly as distant from one another as are sequences from HIV-1 groups M, N, and O, HIV Nomenclature Committee decided to use "groups" rather than "subtypes" for HIV-2 genetic classification. Among these seven HIV-2 groups, only HIV-2 groups A and B are disseminated into significant numbers of human populations (Berry *et al.*, 2001; Schim van der Loeff *et al.*, 1999). HIV-2 group A has been identified predominantly in the western part of West Africa including Guinea, Bissau, Senegal, Gambia, and Mali. In contrast, HIV-2 subtype B has been found in central and eastern West African countries, including Ivory Coast, Ghana, and Nigeria.

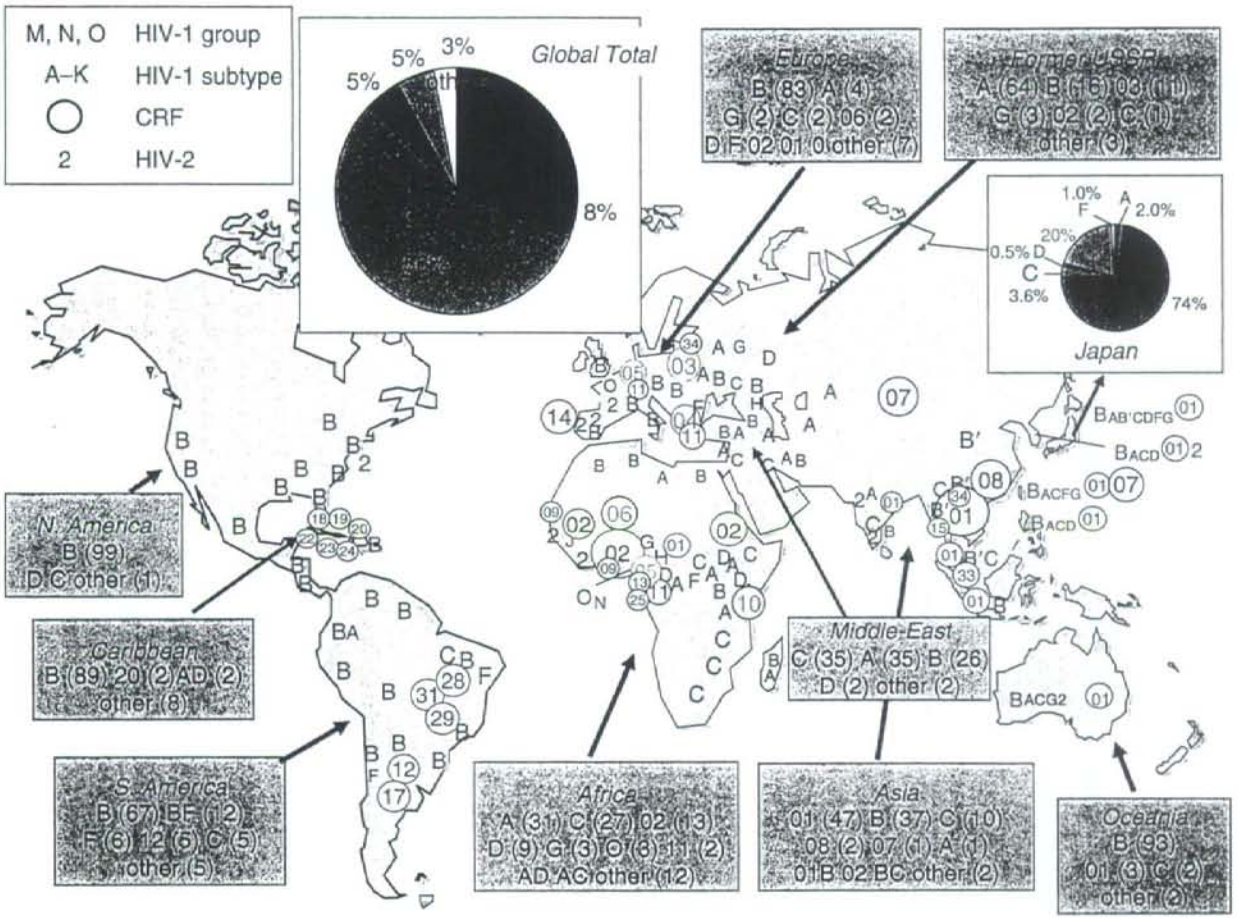

FIGURE 4 Global distribution of HIV genotypes and their estimated proportions. The number in parenthesis after each genotype is the proportion (in percentage) of the indicated genotype in the respective geographic regions. Data source (http://www.hiv.lanl.gov/components/hiv-web/new_geography/). A global total is adopted from Esparza and Bhamarapravati (2000). HIV-1 genotype distribution in Japan is also included. (See Color Plate Section.)

Outside of Africa, unique HIV-2 transmission focuses and clusters were observed in India and Korea. HIV-2 infections were reported in West and South India, frequently associated with dual infections of subtype C strains (Grez *et al.*, 1994; Pfutzner *et al.*, 1992). Ten cases of HIV-2 infections have been reported so far in Korea, where two distinct clusters of HIV-2 (group A) infections were recognized (Nam *et al.*, 2006). The index case (seamen) responsible for each cluster appears to be infected in West Africa and subsequently transmitted the virus to its sexual partners through heterosexual and homosexual contacts inside Korea. While most of the HIV-2 infection cases in Asia are the infections with HIV-2 group A, one HIV-2 group B infection was identified in a person with Korean nationality living in Japan (Kusagawa *et al.*, 2003), who was likely to be infected through heterosexual contacts in DRC (Nam *et al.*, 2006).

IV. Global Distribution of HIV Genotypes

A. Global HIV-1 Variability

On a global scale, the most prevalent HIV-1 genotypes are subtypes C (56%), A (23%), B (8%), D (5%), and CRF01_AE (5%) in 1999 (Esparza and Bhamarapravati, 2000) (Fig. 4). The greatest genetic diversity of HIV-1 has been found in central Sub-Saharan Africa. Subtypes A and C are most common, but all groups and subtypes have been identified. The extensive diversification of HIV-1 group M appears to occur within or near the DRC, where the highest diversity of group M has been recorded (Kalish *et al.*, 2004; Vidal *et al.*, 2000), and the earliest case of HIV-1 infection, dating back to 1959 (Zhu *et al.*, 1998), has been documented. It is consistent to the notion that central Sub-Saharan Africa is the likely origin of current pandemic. In South and East Africa, subtype C predominates (Novitsky *et al.*, 1999) that causes the worst epidemic in those regions with the adult HIV prevalence of more than 30%. In West and West central Africa, the majority of circulating strains is CRF02_AG (Carr *et al.*, 1998). Subtype B viruses remain the most prevalent isolates in North and South America, Western and Central Europe, and Australia, and are also common in several countries in Asia (Hongkong, Japan, Korea, and Taiwan), northern Africa, the Middle East, and among South African and Russian homosexual men. In South America, subtype B is prevalent, while subtypes F and C, and CRF12_BF (Thomson *et al.*, 2000) and other B/F recombinants (CRF17, 28 and 29_BF) (Thomson *et al.*, 2002) have been reported (Section IV.C).

B. HIV-1 Variants in Asia

Studies revealed unique profiles of HIV-1 genotype distribution in Asia. HIV-1 subtype C predominates in India (Lole *et al.*, 1999), with estimated

5.7 million (3.4–9.4 million) people infected. HIV-1 CRF01_AE of central African ancestry is widely circulating in Southeast Asia (Ou *et al.*, 1992, 1993; Weniger and Brown, 1996). CRF01_AE epidemic broke out in the late 1980s among female commercial workers and their clients in Thailand and became prevalent among injecting drug users (IDUs) in this country (Ou *et al.*, 1993) and disseminated to neighboring countries in Asia, including Cambodia, Vietnam, Malyasia, China, Taiwan, Korea, and Japan (Weniger and Brown, 1996; Weniger *et al.*, 1994). Subtype B′ (Thailand variant of subtype B; also referred to as Thai-B virus) (Kalish *et al.*, 1995; Ou *et al.*, 1992, 1993) is a unique subtype B variant that spreads primarily through IDU networks in Southeast Asia (Motomura *et al.*, 2003; Weniger *et al.*, 1994). Two closely related CRFs, CRF07_BC and CRF08_BC, are disseminating rapidly among IDUs in northwestern (Xinjiang Province) (Su *et al.*, 2000) and southeastern (Guangxi Province) China (Piyasirisilp *et al.*, 2000), respectively. A variety of novel CRFs composed of CRF01_AE and subtype B (B′) in Thailand [CRF15_01B (Tovanabutra *et al.*, 2003) and CRF34_01B] and in Malaysia (CRF33_01B) have been identified (Tee *et al.*, 2006).

C. Other HIV-1 Variants of Geographical Relevance

Less prevalent HIV-1 subtypes, but common on a localized scale, are observed in various geographic areas: subtype D, distributed mainly in East Africa (Uganda, Tanzania, and Kenya); subtype F (subsubtype F1) predominant in Romania (mostly children infected through contaminated blood products and unsterilized needle and syringes), and found in a minority of HIV-1-infected people in Brazil; subtype G circulating in West and central Africa, with the highest prevalence in Nigeria as well as in Portugal and northern Spain. A variety of novel CRFs were identified in South America: CRF12_BF (Thomson *et al.*, 2000) and CRF17_BF in Argentina; CRF28_BF, CRF29_BF, and CRF31_BC in Brazil. These new recombinant strains account for ~12% of HIV-1 infections in Latin America (Fig. 4). In Europe, CRF14_BG and its related recombinants are circulating locally in northwestern Spain (Delgado *et al.*, 2002; Thomson *et al.*, 2001) and Portugal (Esteves *et al.*, 2002). The widest range of novel CRFs, including CRF18_cpx (Thomson *et al.*, 2005), CRF19_cpx (Casado *et al.*, 2005), CRF20_BG, CRF22_01A1, CRF23_BG, and CRF24_BG (http://hiv-web.lanl.gov/CRFs/), have been reported in Cuba, where those recombinants account for ~20% of HIV-1 infections (http://www.hiv.lanl.gov/components/hiv-web/new_geography/). Injecting drug use triggered a new HIV-1 epidemic in Eastern Europe: CRF03_AB was originally identified among IDUs in the Russian city of Kaliningrad (Liitsola *et al.*, 1998), and later detected in several cities in Ukraine and Belarus (St. Petersburg, Smolensk, and Perm) (Figs. 3 and 4). Other minor nonrecombinant subtypes (A2, F2, H, J, and K) were detected in central Africa. Most of the remaining CRFs have lesser relevance in epidemic on a global scale.

D. Emergence of HIV-1 Recombinants Worldwide

Although the exact prevalence of recombinant strains is not well known, the preliminary data show that the proportions of discordant *gag/env* samples varied from less than 10% to up to 40% in Africa (McCutchan *et al.*, 1999; Vidal *et al.*, 2000) and 10–30% in some areas in Asia, including central Myanmar (Kusagawa *et al.*, 1998; Motomura *et al.*, 2000, 2003; Takebe *et al.*, 2003) and more than 60% in western part of Yunnan Province (southwestern China) (Yang *et al.*, 2002).

Recombinant viruses have already contributed substantially to the global pandemic, and the likelihood of generating recombinant viruses will continue to increase as the different HIV-1 subtypes spread worldwide (Peeters, 2000). Mixing of different lineages of HIV-1 strains could quickly lead to the evolution of new recombinant strains. Even recombinant viruses will recombine, leading to the evolution of second-generation recombinants, inter-CRF recombinants (ICRs): ICR01_0708 identified among IDUs in Yunnan Province of China is the first example of this category, which is composed of two closely related CRFs in China, CRF07_BC and CRF08_BC (Yang *et al.*, 2003) (Fig. 3).

V. Methods for Identifying HIV Genetic Forms

A. Phylogenetic Sequence Analysis

For subtype classification, phylogenetic sequence analysis is the most reliable method. Various software programs for phylogenetic analysis (i.e., molecular evolutionary genetic analysis, MEGA) are freely available (http://hiv-web.lanl.gov/). Due to the high frequency of recombination in HIV, it is necessary to equip software programs designed for identifying recombination. This is particularly the case for the molecular epidemiological investigation in the areas where different lineages of HIV-1 strains are cocirculating. Softwares designed for detecting recombination, including Simplot (Ray, 2002) and Recombination Identification Program (RIP) (http://hiv-web.lanl.gov/) are useful for this purpose. Figure 5 illustrates an example of phylogenetic tree analysis and recombination breakpoint analyses (bootscanning plot and subregion tree analyses) for identification and characterization of novel recombinant strains in Myanmar (Takebe *et al.*, 2003).

B. Alternative Methods (Heteroduplex Mobility Assay and Serotyping)

Other methods, less expensive and requiring less sophisticated equipments, can be useful as the alternatives for sequencing. This includes serotyping and heteroduplex mobility assay (HMA) (Delwart *et al.*, 1993).

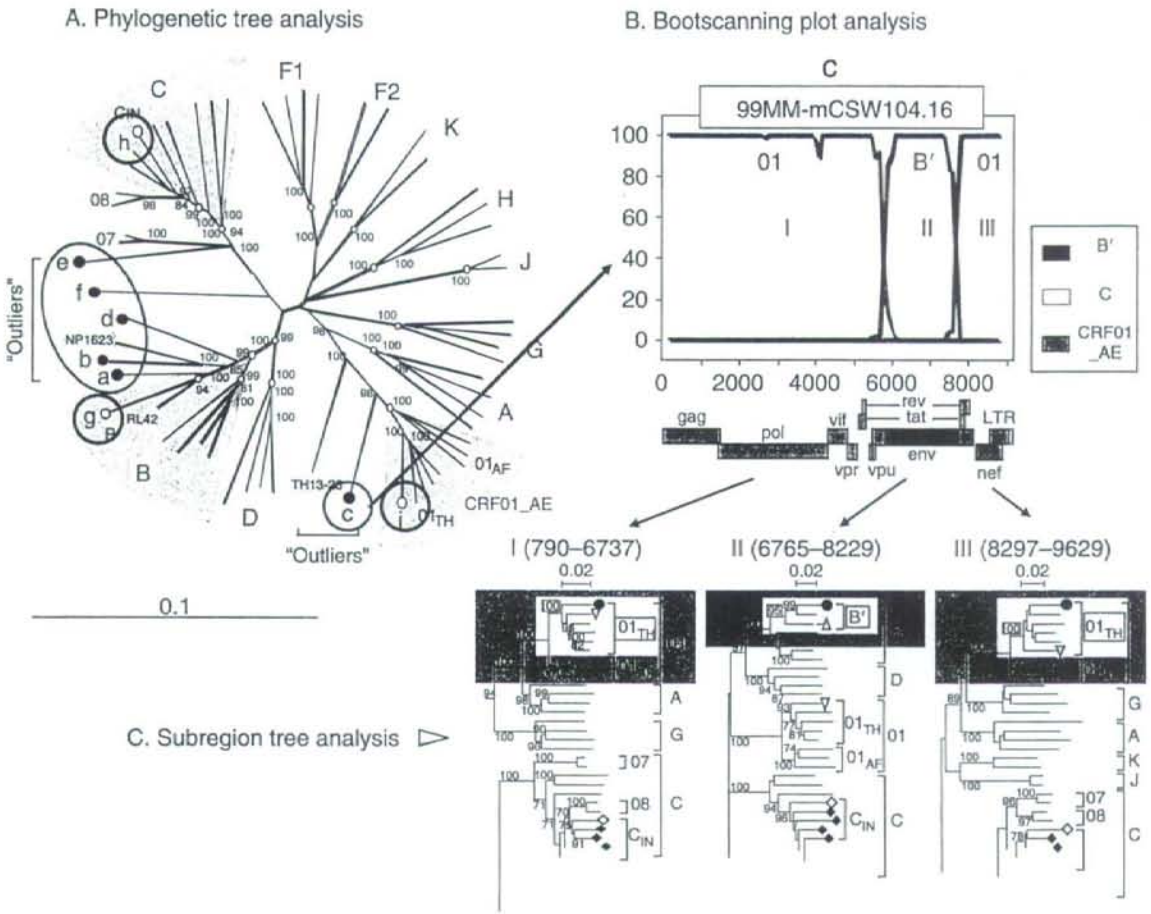

FIGURE 5 Phylogenetic tree analysis (A) and recombination breakpoint analysis for identification novel HIV-1 recombinant strains. Neighbor-joining tree analysis of Myanmar HIV-1 isolates based on near full-length sequences. Strains (g–i) belong to nonrecombinant forms of HIV-1 subtypes B′

Serotyping is a method based on the binding of antibodies present in the patient's sera to the peptides representing a segment of envelop V3 loop of different subtypes (Pau *et al.*, 1993, 1994). Serotyping is particularly useful for analyzing large numbers of specimens for epidemiological studies. However, this assay cannot reliably distinguish between subtypes A and C and cannot detect recombinants. HMA is the method based on electrophoretic mobility of DNA duplexes formed by hybridization of the polymerase chain reaction (PCR)-amplified sequences with reference sequences of different subtypes. However, both methods are less reliable in the areas with high HIV-1 genetic heterogeneity, such as central Africa and some regions in Asia, because of the high frequency of intermediate or incorrect results. These two methods can be useful in the areas, where one or two relatively homogeneous subtypes are prevalent. For instance, serotyping method was successfully applied for the distinction of subtype B and CRF01_AE infections for the study in Thailand (Pau *et al.*, 1993).

VI. Origin of HIVs and Genesis of HIV-1 Pandemic

A. HIV/AIDS as a "Zoonosis"

Current evidence indicates that HIV-1 and HIV-2 entered into human population through multiple zoonotic infections from SIVs-infected nonhuman primates (Hahn *et al.*, 2000). It has been known that HIV-2 and SIV sm have a high degree of genetic and phenotypic homology (Gao *et al.*, 1992), sharing unique open-reading frame, called vpx, in their genomes. Moreover, the habitat of the sooty mangabey closely matches HIV-2 endemicity in West Africa. These close relationships between HIV-2 and SIV-sm led to the hypothesis that HIV-2 infection is a zoonosis (Sharp *et al.*, 1999).

By contrast, HIV-1 is most closely related to SIV cpz isolated from the chimpanzee subspecies *Pantroglodytes troglodytes* (*P.t.t.*) (Corbet *et al.*, 2000; Gao *et al.*, 1999; Hahn *et al.*, 2000; Peeters *et al.*, 1997). The most diverse forms of HIV-1 are found in the geographic region corresponding to the range of *P.t.t.* in West equatorial Africa (Charneau *et al.*, 1994; De Leys *et al.*, 1990; Gurtler, 1996; Simon *et al.*, 1998), and HIV-1 groups and

(Thailand variant of subtype B) and C, and CRF01_AE, respectively. Strains (a–f) are "outliers" that are not assigned to any known HIV-1 genotypes (subtypes/CRFs). They turned out to be novel HIV-1 intergenotype recombinants, that is, "unique recombinant forms" (URFs). The data outputs obtained from various recombination breakpoint analyses, including bootscanning plot analysis (B) and subregion tree analysis (C) for "outlier" strain (c) are schematically illustrated. The results indicate that the strain (c) is a novel HIV-1 URF composed of subtype B′ and CRF01_AE of Thailand origins. The strain (c) shows the structural similarity to CRF15_01B, but is not exactly identical (Takebe *et al.*, 2003). (See Color Plate Section.)

SIV cpz sequences are interspersed in phylogenetic trees, suggesting that there are shared viral lineages in human and chimpanzees (Corbet *et al.*, 2000; Gao *et al.*, 1999; Hahn *et al.*, 2000; Peeters *et al.*, 1997). Each group of HIV-1 and HIV-2 is believed to represent a distinct cross-species transmission of the viruses from its chimpanzee and sooty mangabey reservoirs, respectively (Hahn *et al.*, 2000). However, genetic survey of SIVs in African primate species in central Africa using fecal specimens identified HIV-1 group O-like viruses in wild gorilla (*Gorilla gorilla gorilla, G.g.g.*) (Van Heuverswyn *et al.*, 2006). Collectively, it could be speculated that SIV-cpz*Ptt* have crossed at least twice in humans, resulting in the AIDS pandemic by HIV-1 group M in one instance and infection of a few individuals in Cameroon by group N in another, and that the third HIV-1 lineage group O appears to be evolved from a virus from wild gorilla (*G.g.g*), while this virus (SIV gor) forms monophyletic lineages within SIV-cpz*Ptt* radiation (Sharp *et al.*, 2005). The plausible origins and routes of cross-species transmissions of HIVs are illustrated in the left side of Fig. 2.

B. Dating the Origin of Pandemic HIV-1 Strains

Korber *et al.* (2000) estimated the date of the most recent common ancestor (MRCA) of HIV-1 group M to be 1931 [95% confidence interval (CI): 1915–1941], suggesting that HIV-1 group M began its expansion in human population roughly 70 years ago. The phylogenetic analyses assuming molecular clock suggested that the founder of subtype B in the United States originated in 1967 (95% CI: 1960–1971). Similarly, the MRCA of CRF01_AE in Thailand was dated 1986 (95% CI: 1978–1989) (Korber *et al.*, 2000). By contrast, according to Lemey *et al.* (2003), the date of the MRCA of HIV-2 group A strains was estimated to be 1940 $\pm$ 16, and that of HIV-2 group B strains was estimated to be 1945 $\pm$ 14 in Guinea, Bissau. Taken together, zoonotic transfers of HIVs occurred in early or the first-half of the twentieth century and subsequently spread globally, generating the pandemic observed today. The origins and plausible route of dissemination of HIV-1 strains responsible for epidemic in Asia is schematically illustrated in Fig. 6.

VII. Biological Significance of HIV-1 Variability and Recombination

A. HIV-1 Subtypes and Disease Progression

It has been suggested that HIV-1 subtypes could influence viral transmissibility and pathogenecity. However, the existence of many other factors makes it difficult to establish the true effect of viral subtypes. A study in Thailand showed that the disease progression in the patients infected with

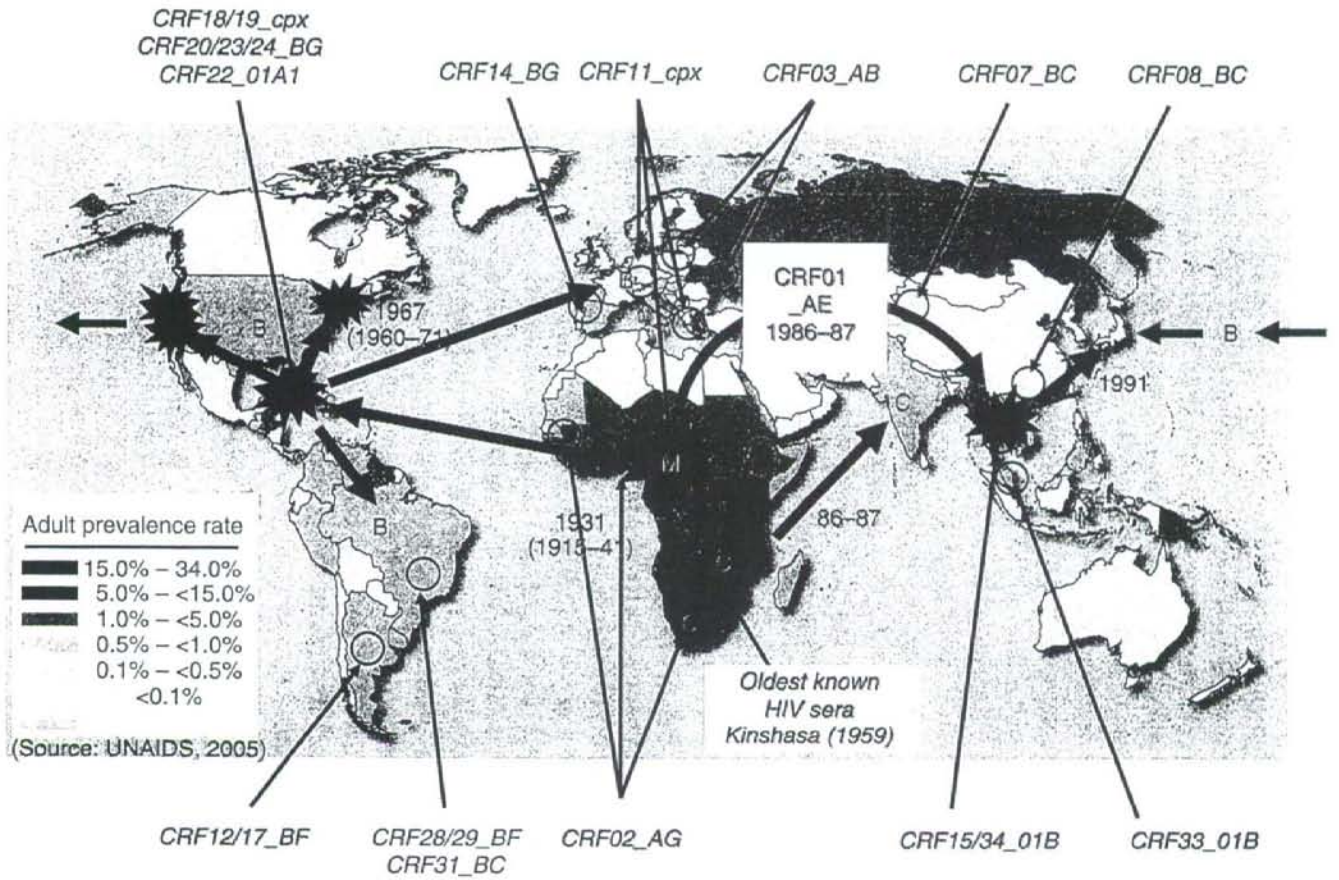

FIGURE 6 Origin of HIV-1 group M and plausible routes of the spread of the HIV-1 strains responsible for epidemic in Asia. The epidemic focuses of selected CRFs of geographical relevance are shown. The illustrations are superimposed on the world map with estimated adult HIV prevalence in different countries (UNAIDS/WHO, 2005). (See Color Plate Section.)

CRF01_AE is similar to those observed in subtype B-infected populations in the West (Amornkul *et al.*, 1999; Kilmarx *et al.*, 2000).

In contrast, several studies showed that HIV-1 subtypes differ in rates of progression to AIDS: (1) A prospective study of female prostitutes in Senegal showed that women infected with subtypes C, D, or G were eightfold more likely to develop AIDS than subtype A (Kanki *et al.*, 1999); (2) The cohort study in Kenya, where subtypes A, C, and D cocirculate, plasma RNA levels were found to be highest in subtype C (Neilson *et al.*, 1999); (3) A study in Tanzania suggests that subtypes A and C and recombinant viruses are more likely to be transmitted perinatally than subtype D (Renjifo *et al.*, 2001), suggesting that maternal subtype may play a role in vertical transmission; (4) The response to the proinflammatory cytokine tumor necrosis factor alpha (TNF-α) is increased in subtype C long terminal repeat (LTR) with nuclear factor κB (NFκB) configuration, suggesting that subtype C may have a replication advantage in individuals with chronic immune activation (Montano *et al.*, 2000); (5) A matched case-control study showed that viruses containing subtype C LTRs were six times more likely to be transmitted than those with subtype D (Blackard *et al.*, 2001). The study in Uganda showed subtype D was associated with faster progression to death and with a lower CD4 cell count than subtype A (Kaleebu *et al.*, 2002).

In contrast, a study from Sweden showed no differences in disease progression in subtypes A, B, C, or D (Alaeus *et al.*, 1999). However, it is not clear whether such differences are due to the environmental factors such as the prevalence of other infectious diseases that may induce the systemic "immune activation," including sexually transmitted diseases (STDs) and parasitic diseases. Indeed, several studies suggest faster disease progression in the persons infected in Africa, compared with those infected in the United States or Western Europe (Galai *et al.*, 1997; Kanki *et al.*, 1999). Long-term prospective studies in recent seroconverters will be needed to elucidate the relationship between HIV-1 genotypes and clinical disease progression.

B. HIV-1 Dual Infection, Superinfection, and Recombination

1. Dual Infection: Mechanism and Prevalence

Dual infections are the prerequisite for the generation of recombinants. When a single cell that is infected with genetically distinct viruses produces progeny virions with RNAs from each virus, recombination could occur between the two copackaged heterologous RNAs through strand switching during the next replication cycle (Malim and Emerman, 2001). Therefore, dual infection with more than one lineage of HIV-1 strains within an individual is a source of rapid viral evolution by recombination. Over the last decade, a number of cases of dual infections with the same or different HIV-1 subtypes through various transmission routes, including vertical

transmission (Janini *et al.*, 1998; Mellquist *et al.*, 1999), sexual transmission (Jost *et al.*, 2002; Zhu *et al.*, 1995), and blood transfusion or injection drug use (Diaz *et al.*, 1995; Ramos *et al.*, 2002; Sala *et al.*, 1994, 1995) have been documented.

2. Distinction Between Coinfection and Superinfection

By the temporal mode of the acquisitions of different HIV-1 strains, dual infections are divided into two categories: coinfection (simultaneous) and superinfection (sequential). Coinfection is defined as an infection with two heterologous strains ether simultaneously or within a brief period of time (arbitrarily within the first month of infection) before infection with the first strain has been established and an immune response has developed. In contrast, superinfection is defined as an infection with a second strain after the immune response to the first infection has been established (Smith *et al.*, 2005). As of August 2005, 16 published cases of superinfections have been reported worldwide (Smith *et al.*, 2005). In several reported cases, superinfection has resulted in recombination between the initial and the secondary strains (Fang *et al.*, 2004; McCutchan *et al.*, 2005; Yang *et al.*, 2005).

3. Superinfection: Implications for Vaccine Development

The majority of superinfection appears to has occurred in the early stage of infection. In contrast, several population-based studies reported the rarity of superinfection during chronic infections (Gonzales *et al.*, 2003; Tsui *et al.*, 2004). The reported rarity of superinfection during chronic HIV-1 infection may reflect the time required for the immune responses to mature and may suggest that immune responses in the infected host could protect against superinfection. This may offer hope for an effective vaccine against HIV-1. However, results from several published studies appear to indicate that even strong $CD8^+$ T-cell-mediated responses against the initial infection may not be sufficient for the protection against superinfection (Altfeld *et al.*, 2002; Jost *et al.*, 2002; Yang *et al.*, 2005). Moreover, in most reported cases of superinfection, patients have experienced a decrease in $CD4^+$ cell count and increase in HIV load, accelerating disease progression (Gottlieb *et al.*, 2004; Jost *et al.*, 2002; Smith *et al.*, 2005). The knowledge of superinfection is thus vital to understand the changes in viral pathogenesis and the host immunity and provides the important insights into future vaccine strategies.

C. Biological Implications of Recombination

Recombinant viruses may have certain advantages over the parental strain, including modifications in tropism and replication efficiency ("viral fitness"). Under selection pressure imposed by antiretroviral drugs, recombination between strains with different drug sensitivity resulted in new HIV-1 variants with dual or multiple drug resistance (Moutouh *et al.*, 1996). The

discovery of large numbers of recombinant strains clearly suggests that coinfection with different HIV-1 strains is not rare as once thought. The dual infections with different subtypes have been reported in the regions where multiple variants are cocirculating. Furthermore, as described in the previous section, a study showed that HIV-1 superinfection can occur in the setting of a strong and broadly directed virus-specific $CD8^+$ T-cell response (Altfeld *et al.*, 2002; Jost *et al.*, 2002), suggesting that the host immunological responses are not efficient against divergent strains. These findings would provide important implications for vaccine development as well as for the prevention efforts from public health viewpoints.

VIII. Conclusions

Global dissemination of HIVs represents a dramatic and deadly example of recent genome emergence and expansion. As reviewed in this chapter, recent studies revealed that a pandemic HIV strain, HIV-1 group M, began its expansion in human population roughly 70 years ago (early twentieth century), it has been diversifying rapidly, now comprising a number of different subtypes and CRFs, and that new recombinant strains are arising continually, becoming a powerful force in global HIV-1 spread. Studies also provide information to delineate the mechanism of viral evolution and for the studies on biological features of HIV strains related to pathogenecity and disease progression. However, the biological significance of the global diversity of HIV-1 strains remains to be defined. Although the immune correlates for protection are still incompletely understood, the extensive variation of HIVs could probably be important in the formulation of the vaccine immunogens. In conclusion, molecular epidemiological information on the HIV strains is critically important to elucidate the dynamics of HIV spread and to formulate future vaccine and other prevention strategies.

Acknowledgments

We thank Tee Kok Keng, Shigeru Kusagawa, and Midori Kawasaki for assistance with the manuscript and references, and Drs. Kuan-Teh Jeang and Naoki Yamamoto for their encouragement and advice. We acknowledge grant support from Ministry of Health, Labour and Welfare, Ministry of Education, Science and Technology in Japan and Japanese Foundation for AIDS Prevention. We also thank our collaborators and colleagues in Asia.

References

Alaeus, A., Lidman, K., Bjorkman, A., Giesecke, J., and Albert, J. (1999). Similar rate of disease progression among individuals infected with HIV-1 genetic subtypes A-D. *AIDS* **13**, 901–907.

Altfeld, M., Allen, T. M., Yu, X. G., Johnston, M. N., Agrawal, D., Korber, B. T., Montefiori, D. C., O'Connor, D. H., Davis, B. T., Lee, P. K., Maier, E. L., Harlow, J., *et al.* (2002). HIV-1 superinfection despite broad CD8+ T-cell responses containing replication of the primary virus. *Nature* 420, 434–439.

Amornkul, P. N., Tansuphasawadikul, S., Limpakarnjanarat, K., Likanonsakul, S., Young, N., Eampokalap, B., Kaewkungwal, J., Naiwatanakul, T., Von Bargen, J., Hu, D. J., and Mastro, T. D. (1999). Clinical disease associated with HIV-1 subtype B′ and E infection among 2104 patients in Thailand. *AIDS* 13, 1963–1969.

Berry, N., Ariyoshi, K., Balfe, P., Tedder, R., and Whittle, H. (2001). Sequence specificity of the human immunodeficiency virus type 2 (HIV-2) long terminal repeat 3 region *in vivo* allows subtyping of the principal HIV-2 viral subtypes A and B. *AIDS Res. Hum. Retroviruses* 17, 263–267.

Blackard, J. T., Renjifo, B., Fawzi, W., Hertzmark, E., Msamanga, G., Mwakagile, D., Hunter, D., Spiegelman, D., Sharghi, N., Kagoma, C., and Essex, M. (2001). HIV-1 LTR subtype and perinatal transmission. *Virology* 287, 261–265.

Carr, J. K., Salminen, M. O., Albert, J., Sanders-Buell, E., Gotte, D., Birx, D. L., and McCutchan, F. E. (1998). Full genome sequences of human immunodeficiency virus type 1 subtypes G and A/G intersubtype recombinants. *Virology* 247, 22–31.

Casado, G., Thomson, M. M., Sierra, M., and Najera, R. (2005). Identification of a novel HIV-1 circulating ADG intersubtype recombinant form (CRF19_cpx) in Cuba. *J. Acquir. Immune Defic. Syndr.* 40, 532–537.

Charneau, P., Borman, A. M., Quillent, C., Guetard, D., Chamaret, S., Cohen, J., Remy, G., Montagnier, L., and Clavel, F. (1994). Isolation and envelope sequence of a highly divergent HIV-1 isolate: Definition of a new HIV-1 group. *Virology* 205, 247–253.

Cilla, G., Rodes, B., Perez-Trallero, E., Arrizabalaga, J., and Soriano, V. (2001). Molecular evidence of homosexual transmission of HIV type 2 in Spain. *AIDS Res. Hum. Retroviruses* 17, 417–422.

Corbet, S., Muller-Trutwin, M. C., Versmisse, P., Delarue, S., Ayouba, A., Lewis, J., Brunak, S., Martin, P., Brun-Vezinet, F., Simon, F., Barre-Sinoussi, F., and Mauclere, P. (2000). *env* sequences of simian immunodeficiency viruses from chimpanzees in Cameroon are strongly related to those of human immunodeficiency virus group N from the same geographic area. *J. Virol.* 74, 529–534.

Cuevas, M. T., Ruibal, I., Villahermosa, M. L., Diaz, H., Delgado, E., Parga, E. V., Perez-Alvarez, L., de Armas, M. B., Cuevas, L., Medrano, L., Noa, E., Osmanov, S., *et al.* (2002). High HIV-1 genetic diversity in Cuba. *AIDS* 16, 1643–1653.

Damond, F., Worobey, M., Campa, P., Farfara, I., Colin, G., Matheron, S., Brun-Vezinet, F., Robertson, D. L., and Simon, F. (2004). Identification of a highly divergent HIV type 2 and proposal for a change in HIV type 2 classification. *AIDS Res. Hum. Retroviruses* 20, 666–672.

De Leys, R., Vanderborght, B., Vanden Haesevelde, M., Heyndrickx, L., van Geel, A., Wauters, C., Bernaerts, R., Saman, E., Nijs, P., Willems, B., Taelman, H., van der Groen, G., *et al.* (1990). Isolation and partial characterization of an unusual human immunodeficiency retrovirus from two persons of west-central African origin. *J. Virol.* 64, 1207–1216.

Delgado, E., Thomson, M. M., Villahermosa, M. L., Sierra, M., Ocampo, A., Miralles, C., Rodriguez-Perez, R., Diz-Aren, J., Ojea-de Castro, R., Losada, E., Cuevas, M. T., Vázquez-de, E., *et al.* (2002). Identification of a newly characterized HIV-1 BG intersubtype circulating recombinant form in Galicia, Spain, which exhibits a pseudotype-like virion structure. *J. Acquir. Immune Defic. Syndr.* 29, 536–543.

Delwart, E. L., Shpaer, E. G., Louwagie, J., McCutchan, F. E., Grez, M., Rubsamen-Waigmann, H., and Mullins, J. I. (1993). Genetic relationships determined by a DNA heteroduplex mobility assay: Analysis of HIV-1 env genes. *Science* 262, 1257–1261.

Diaz, R. S., Sabino, E. C., Mayer, A., Mosley, J. W., and Busch, M. P. (1995). Dual human immunodeficiency virus type 1 infection and recombination in a dually exposed transfusion recipient. The Transfusion Safety Study Group. *J. Virol.* **69**, 3273–3281.

Esparza, J., and Bhamarapravati, N. (2000). Accelerating the development and future availability of HIV-1 vaccines: Why, when, where, and how? *Lancet* **355**, 2061–2066.

Esteves, A., Parreira, R., Venenno, T., Franco, M., Piedade, J., Germano De Sousa, J., and Canas-Ferreira, W. F. (2002). Molecular epidemiology of HIV type 1 infection in Portugal: High prevalence of non-B subtypes. *AIDS Res. Hum. Retroviruses* **18**, 313–325.

Fang, G., Weiser, B., Kuiken, C., Philpott, S. M., Rowland-Jones, S., Plummer, F., Kimani, J., Shi, B., Kaul, R., Bwayo, J., Anzala, O., and Burger, H. (2004). Recombination following superinfection by HIV-1. *AIDS* **18**, 153–159.

Galai, N., Kalinkovich, A., Burstein, R., Vlahov, D., and Bentwich, Z. (1997). African HIV-1 subtype C and rate of progression among Ethiopian immigrants in Israel. *Lancet* **349**, 180–181.

Gao, F., Yue, L., White, A. T., Pappas, P. G., Barchue, J., Hanson, A. P., Greene, B. M., Sharp, P. M., Shaw, G. M., and Hahn, B. H. (1992). Human infection by genetically diverse SIVSM-related HIV-2 in west Africa. *Nature* **358**, 495–499.

Gao, F., Bailes, E., Robertson, D. L., Chen, Y., Rodenburg, C. M., Michael, S. F., Cummins, L. B., Arthur, L. O., Peeters, M., Shaw, G. M., Sharp, P. M., and Hahn, B. H. (1999). Origin of HIV-1 in the chimpanzee Pan troglodytes troglodytes. *Nature* **397**, 436–441.

Gonzales, M. J., Delwart, E., Rhee, S. Y., Tsui, R., Zolopa, A. R., Taylor, J., and Shafer, R. W. (2003). Lack of detectable human immunodeficiency virus type 1 superinfection during 1072 person-years of observation. *J. Infect. Dis.* **188**, 397–405.

Gottlieb, G. S., Nickle, D. C., Jensen, M. A., Wong, K. G., Grobler, J., Li, F., Liu, S. L., Rademeyer, C., Learn, G. H., Karim, S. S., Williamson, C., Corey, L., *et al.* (2004). Dual HIV-1 infection associated with rapid disease progression. *Lancet* **363**, 619–622.

Grez, M., Dietrich, U., Balfe, P., von Briesen, H., Maniar, J. K., Mahambre, G., Delwart, E. L., Mullins, J. I., and Rubsamen-Waigmann, H. (1994). 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.* **68**, 2161–2168.

Gurtler, L. (1996). Difficulties and strategies of HIV diagnosis. *Lancet* **348**, 176–179.

Gurtler, L. G., Hauser, P. H., Eberle, J., von Brunn, A., Knapp, S., Zekeng, L., Tsague, J. M., and Kaptue, L. (1994). A new subtype of human immunodeficiency virus type 1 (MVP-5180) from Cameroon. *J. Virol.* **68**, 1581–1585.

Hahn, B. H., Shaw, G. M., De Cock, K. M., and Sharp, P. M. (2000). AIDS as a zoonosis: Scientific and public health implications. *Science* **287**, 607–614.

Hoelscher, M., Kim, B., Maboko, L., Mhalu, F., von Sonnenburg, F., Birx, D. L., and McCutchan, F. E. (2001). High proportion of unrelated HIV-1 intersubtype recombinants in the Mbeya region of southwest Tanzania. *AIDS* **15**, 1461–1470.

Janini, L. M., Tanuri, A., Schechter, M., Peralta, J. M., Vicente, A. C., Dela Torre, N., Pieniazek, N. J., Luo, C. C., Ramos, A., Soriano, V., Schochetman, G., Rayfield, M. A., *et al.* (1998). Horizontal and vertical transmission of human immunodeficiency virus type 1 dual infections caused by viruses of subtypes B and C. *J. Infect. Dis.* **177**, 227–231.

Jost, S., Bernard, M. C., Kaiser, L., Yerly, S., Hirschel, B., Samri, A., Autran, B., Goh, L. E., and Perrin, L. (2002). A patient with HIV-1 superinfection. *N. Engl. J. Med.* **347**, 731–736.

Kaleebu, P., French, N., Mahe, C., Yirrell, D., Watera, C., Lyagoba, F., Nakiyingi, J., Rutebemberwa, A., Morgan, D., Weber, J., Gilks, C., and Whitworth, J. (2002). Effect of human immunodeficiency virus (HIV) type 1 envelope subtypes A and D on disease progression in a large cohort of HIV-1-positive persons in Uganda. *J. Infect. Dis.* **185**, 1244–1250.

Kalish, M. L., Baldwin, A., Raktham, S., Wasi, C., Luo, C. C., Schochetman, G., Mastro, T. D., Young, N., Vanichseni, S., Rubsamen-Waigmann, H., von Briesen, H., Mullins, E., *et al.*