

CHAPTER 2

LITERATURE REVIEWS

2.1 Biology of HIV

2.1.1 Classification and morphology

HIV was classified in family *Retroviridae*, and subfamily *lentivirinae*. The diameters of virus particle are approximately 110 nm with an outer envelope protein. The actual arrangement is usually in the form of icosahedrons. The outer envelop consists of lipid bilayers membrane derived from the infected cell and glycoprotein and some cellular proteins. The spikes of cell surface consist of a transmembrane glycoprotein, gp41, as the stem and a surface glycoprotein, gp120. (28, 29) Deep inside of the viral envelop is a layer of matrix protein (MA) or p17, in which a conical-shape protein core containing the two identical single stranded plus sense RNA is located. The most protein in the capsid core is the viral capsid protein (CA), or p24.

The genomic RNA tensely associated with the nucleocapsid protein (NC) or p7, and the three enzymes: protease (PR), or p11, reverse transcriptase (RT) or p66/p51, and integrase (IN), or p31. Other viral proteins, such as Vif, Vpr, and Nef, are also found in the virion. The structure of HIV-1 is shown in Figure 2.1.

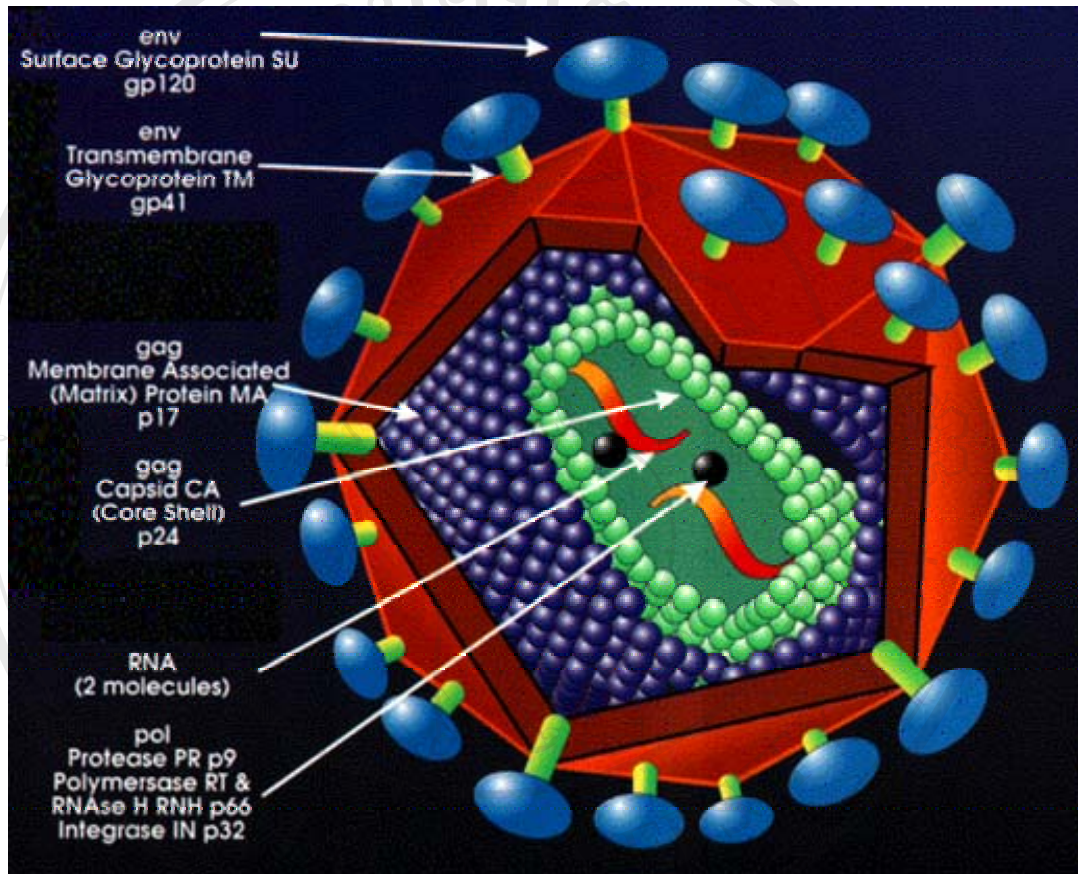


Figure 2.1 The structure of Human Immunodeficiency Virus (HIV)

(Source: [www.wellesley.edu/.../ Chem101/hiv/HIV-1.html](http://www.wellesley.edu/.../Chem101/hiv/HIV-1.html))

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright © by Chiang Mai University
All rights reserved

2.1.2 Genomic structure

HIV genome is 9.7 kilobases (kb) in length. (30) There are three structural genes, *gag*, *pol*, and *env* located from 5' to 3' (31, 32) In addition to the three major genes, HIV viruses also contain several small genes including LTR, *vif*, *vpr*, *tat*, *rev*, *vpu*, and *nef* that control expression of the viral genome. The genomic structure of HIV-1 is shown in Figure 2.2.

LTR Long terminal repeat, this part genome is identical repeated non-coding sequences designated U3, R, and U5. The functions of these units are important for transcription initiation and polyadenylation.

GAG Group specific antigens, the genomic region encoding the capsid proteins. The precursor is p55 myristylated protein, which is processed to p17 (MAtrix), p24 (CApsid), p7 (NucleoCApsid), and p6 proteins, by the viral protease.

POL The genomic region encoding the viral enzymes protease, reverse transcriptase, and integrase, which are produced as a Gag-pol precursor polyprotein.

ENV The responsibility of this gene is to produce outer envelope. Precursor is *gp160* which is processed to the external glycoprotein *gp120* and the transmembrane glycoprotein *gp41*.

VIF Viral infectivity factor, it is likely that the Vif protein increases the efficiency of cell to cell and person to person spread in HIV infection.

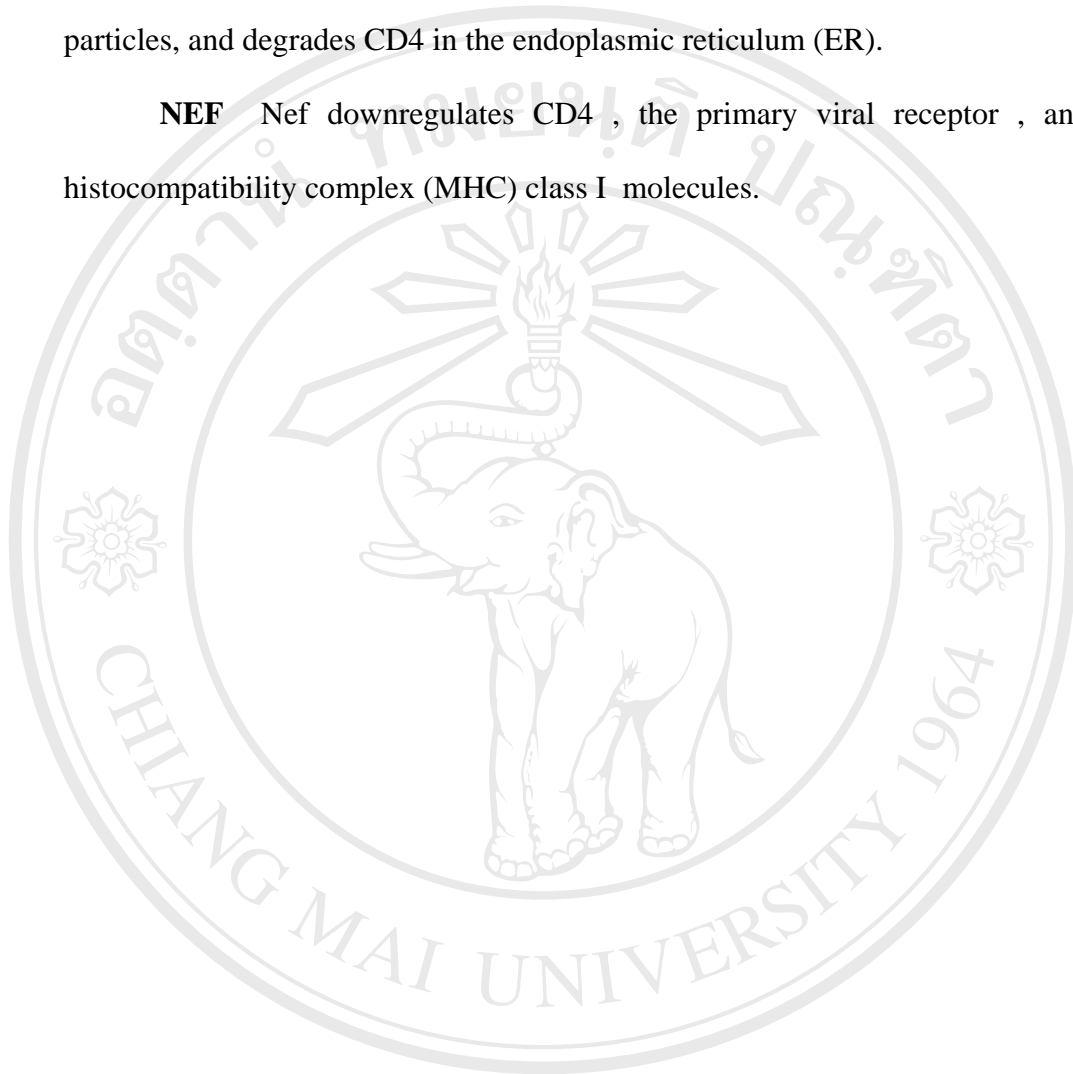
VPR Viral protein R is a protein, which promotes nuclear localization of preintegration complex, inhibits cell division, arrests infected cells at G2/M.

TAT Transactivator of transcription. Tat protein is necessary for viral expression and replication.

REV The second necessary regulatory factor for HIV expression.

VPU Viral protein U is a protein, which promotes extracellular release of viral particles, and degrades CD4 in the endoplasmic reticulum (ER).

NEF Nef downregulates CD4, the primary viral receptor, and major histocompatibility complex (MHC) class I molecules.



ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright © by Chiang Mai University
All rights reserved

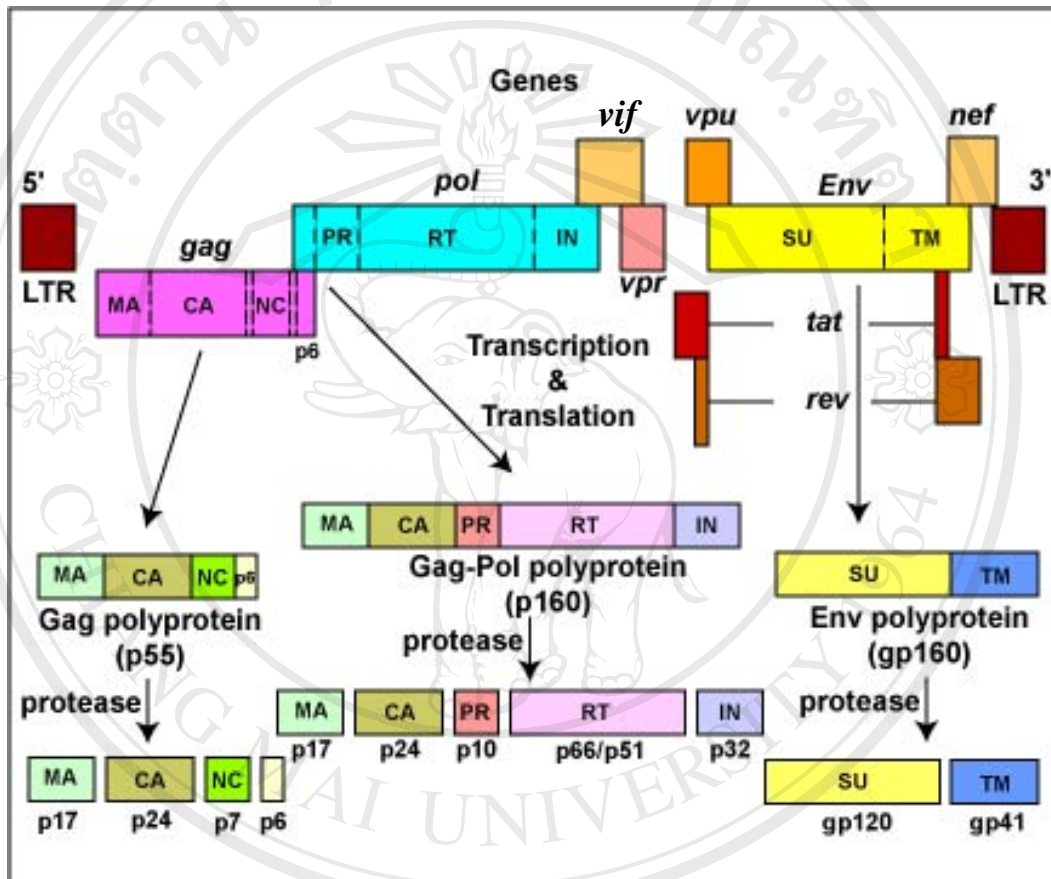


Figure 2.2 HIV-1 genomic structures

(Source: www.cat.cc.md.us)

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
 Copyright © by Chiang Mai University
 All rights reserved

2.1.3 Viral replication

An overview of HIV-1 replication cycle is given in Figure 2.3. For the host cell infection, viral attachment is likely to happen through CD4-gp120 interaction and also with the function of co-receptors to facilitate the viral entry. Once viral entry occurred, the membrane fusion process allows the viral envelop to release of the viral core (uncoating). As a classification in family *Retroviridae*, the replication cycle of HIV-1 requires the conversion of its RNA genome into DNA, so-called reverse transcription using the enzyme reverse transcriptase (RT). (7) Naturally, during viral replication, the reverse transcriptase not only can switch templates, but also lack of proof reading activity. (8) It has been known that viral diversity caused by an error prone RT. (9) Upon double-stranded DNA synthesis, it is subsequently transported to the nucleus, and is integrated into the host cell genome by viral integrase enzyme (IN). The integrated retroviral DNA genome is called the provirus. Next, the proviral DNA is transcribed by the cellular RNA polymerase II into a polyadenylated unspliced mRNA, and consequently translated into the viral proteins. Then, the viral genomic RNA and proteins migrate to the host cell membrane, where they assembled with several host cell proteins, like beta-2 microglobulin, human leukocyte antigens (HLA) to new virions and are still immature (noninfectious form). Finally, they are released by budding. To be infectious particles, the viral protease cleaved the Gag and Gag-Pol precursor proteins into the different subunit proteins.

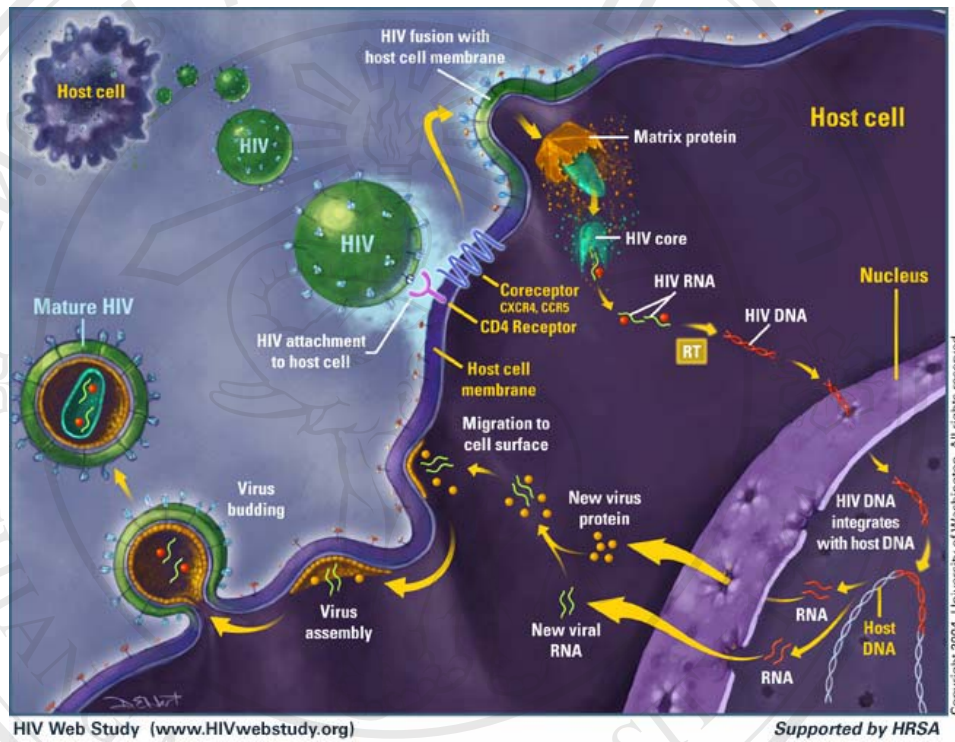


Figure 2.3 HIV-1 replication cycle

(Source: www.HIVwebstudy.org)

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
 Copyright © by Chiang Mai University
 All rights reserved

2.1.4 Genetic diversity of HIV

Acquired immunodeficiency syndrome is caused by infection of HIV-1 and HIV-2. HIV-1 is responsible for the most epidemics (33), while HIV-2 is found mostly in West Africa. (34, 35) HIV-1 strains are genetically diverse. (10, 11, 13, 36). The strains have been diversified extensively through mutations and recombinations since their initial transmission to humans in central Africa. There was a report that the error rate was estimated to be two or three events per genome per round of viral replication. (37). The high error rate of HIV reverse transcriptase in combination with the estimated *in vivo* HIV-1 replication rate of ten billion new virions each day leads to extraordinary genetic diversity of HIV. Within an infected individual the genetic heterogeneity of viral population also occurred and a sub-strain of HIV has been termed a “quasispecies”. (38-42) At present, 9 subtypes and sub-subtypes of the HIV-1 group M are recognized.

In addition, different HIV-1 strains can also recombine at a high rate, generating large genetic diversity. Recombination requires firstly the simultaneous infection of a cell with at least two distinct HIV strains, allowing the encapsulation of one RNA transcript from each provirus into a heterozygous virion. Secondly, the subsequent infection of a new cell, the reverse transcriptase generates a newly synthesized retroviral DNA sequence that is recombinant between the two parental genomes. Finally, a recombinant genome can be generated during reverse transcription. (43) The mechanism of HIV recombination was shown in Figure 2.4.

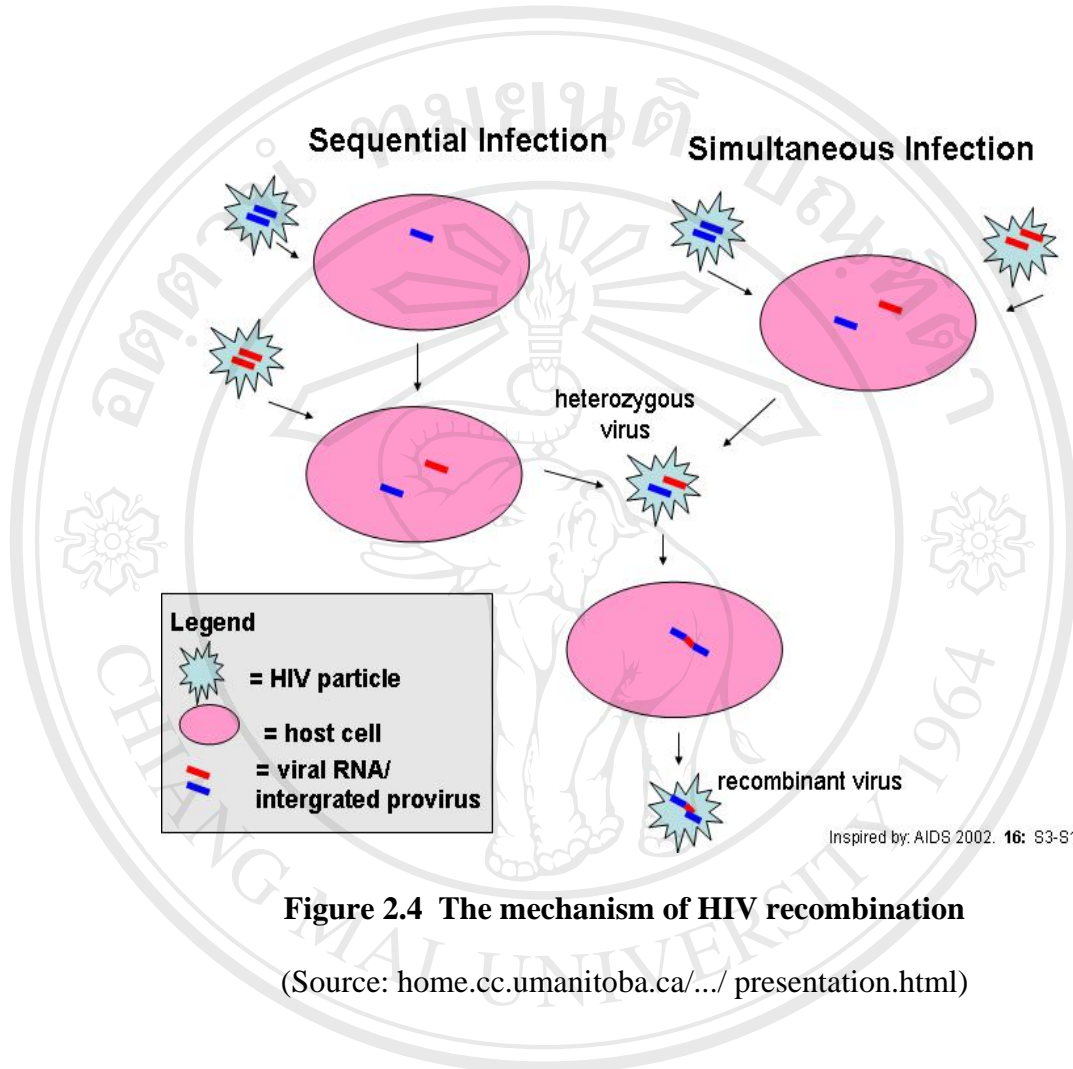


Figure 2.4 The mechanism of HIV recombination

(Source: home.cc.umanitoba.ca/.../presentation.html)

With many evidences of dual infections, both with different subtypes and with different strains of the same subtype, in individuals, (44-48) thirty-three circulating recombinant form (CRF) have been identified at present. These CRFs were established when three or more sequences from epidemiologically unrelated individuals have viruses sharing identical recombinant structures (13) , for example, CRF01_AE (49-51) CRF07_BC(52), CRF10_CD (53) and CRF15_01B. (54) The genomic structures of some circulating recombinant forms (CRF) were shown in Figure 2.5.

2.1.5 HIV nomenclature system

There are four types of categories should be used to refer to the HIV-1 lineage: Group, Subtype, Sub-subtype, and CRF

Group will be used to refer to the very distinctive HIV-1 lineage M (Major), O (Outlier) and N (Non -M, Non-O).

Subtype will be used to refer to the major clades within group M.

Sub-subtype will be used to refer to a distinctive lineage within subtype that is not enough to justify calling a new subtype such as sub-subtype A1 to A4.

Circulating Recombinant Form (CRF) will be used to call a recombinant lineage that occurs oftentimes in the HIV pandemic.

For defining a new subtype, sub-subtype, or CRF, the common rules in HIV nomenclature proposal mention that the representative strains must be identified in at least three individuals with no direct epidemiology related.(13)

Recently, the updated rules, when classifying new sequences, the HIV-1 subtyping set are often used. The criteria for reference sequences should be:

1. Four HIV-1 sequences of each group, subtype, and sub-subtype, are included, if available.
2. The four sequences should coarsely describe the variant of each class as an effective population.
3. Further selection criteria for reference sequences were:
 - a) full length genomes that cover all genes,
 - b) no clear sign of recombinant history,
 - c) published with a peer reviewed citation,
 - d) recent rather than older samples,
 - e) covered major geographic distribution,
 - f) no sign of hypermutation,
 - g) not synthetic, i.e. real sequence from a patient,
 - h) no extreme indels,
 - i) viable and intact as far as known.
4. The CRFs included are now described by one sequence as the prototype. The breakpoint is based on the prototype, and will agree with the updated CRF page.

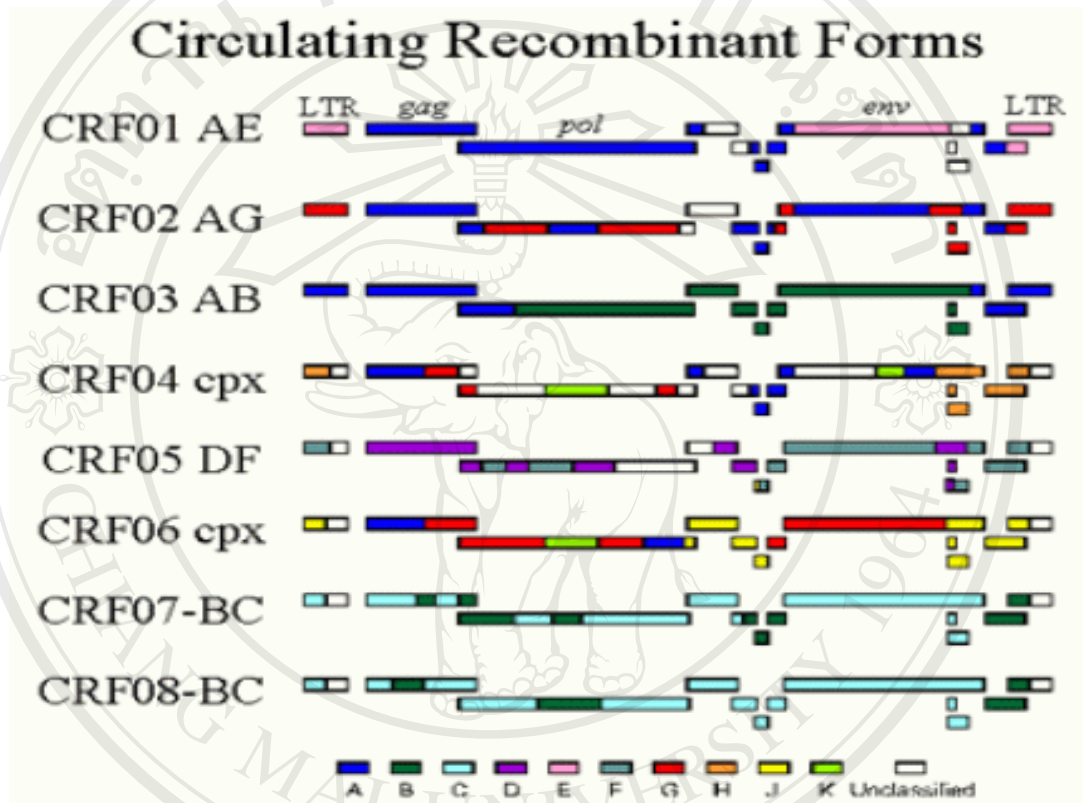


Figure 2.5 An example of Circulating Recombination Form (CRF)

(Source: www.nhrbc.org/HIV_vaccine/paper16.4.html)

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright © by Chiang Mai University
All rights reserved

2.2 Molecular Epidemiology of HIV

2.2.1 The global epidemiology of HIV

Human Immunodeficiency virus (HIV) remains one of the most serious health problems in the world. It is a pathogen which causes Acquired Immunodeficiency Syndrome (AIDS). The global estimates of the people living with HIV/AIDS as of the end of 2004 were 39.4 million adults and children according to the information available to UNAIDS and WHO at the current time. (55) The number of HIV-seroconverters in 2004 was 4.9 million people and almost 50% of newly infected adults were women. The epidemic has not been overcome anywhere. At present, an HIV-1 incident infection remains, and the epidemic is freely out of control in many countries. The global HIV/AIDS epidemic was shown in Table 2.1, and the details of information were showed in Figure 2.6-2.9.

Table 2.1 Global summary of the HIV/AIDS epidemic in December 2004

Number of people living with HIV/AIDS in 2004

Total	39.4 million (35.9 - 44.3 million)
Adults	37.2 million (33.8 - 41.7 million)
Women	17.6 million (16.3 - 19.5 million)
Children under 15 years	2.2 million (2.0 - 2.6 million)

People newly infected with HIV in 2004

Total	4.9 million (4.3 - 6.4 million)
Adults	4.3 million (3.7 - 5.7 million)
Children under 15 years	640,000 (570,000 – 750,000)

AIDS deaths in 2004

Total	3.1 million (2.8 - 3.5 million)
Adults	2.6 million (2.3 - 2.9 million)
Children under 15 years	510,000 (460,000 – 600,000)

(Source: <http://www.unaids.org/wad2004>)

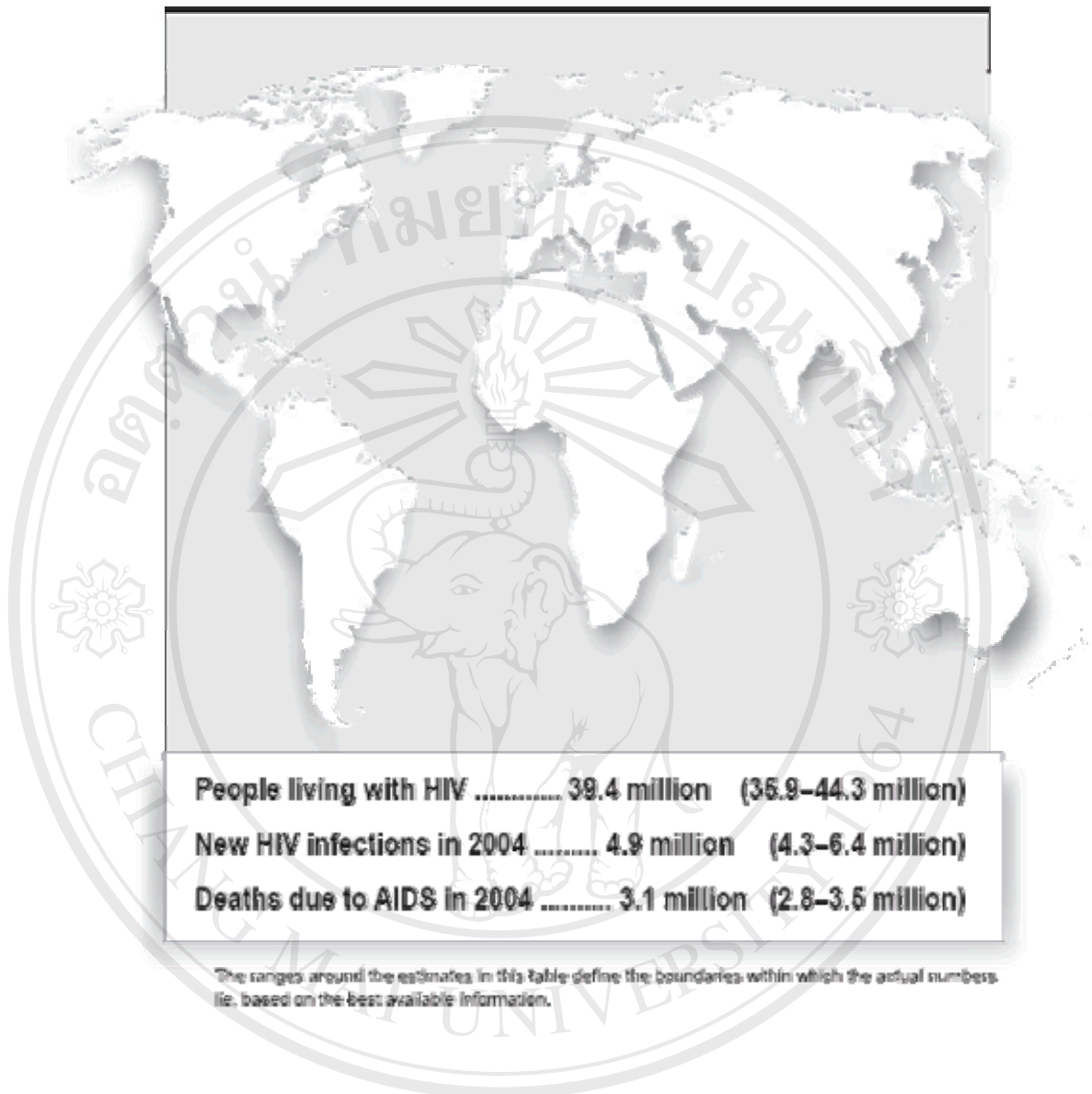


Figure 2.6 Global estimates for adults and children, end 2004

(Source: <http://www.unaids.org>)

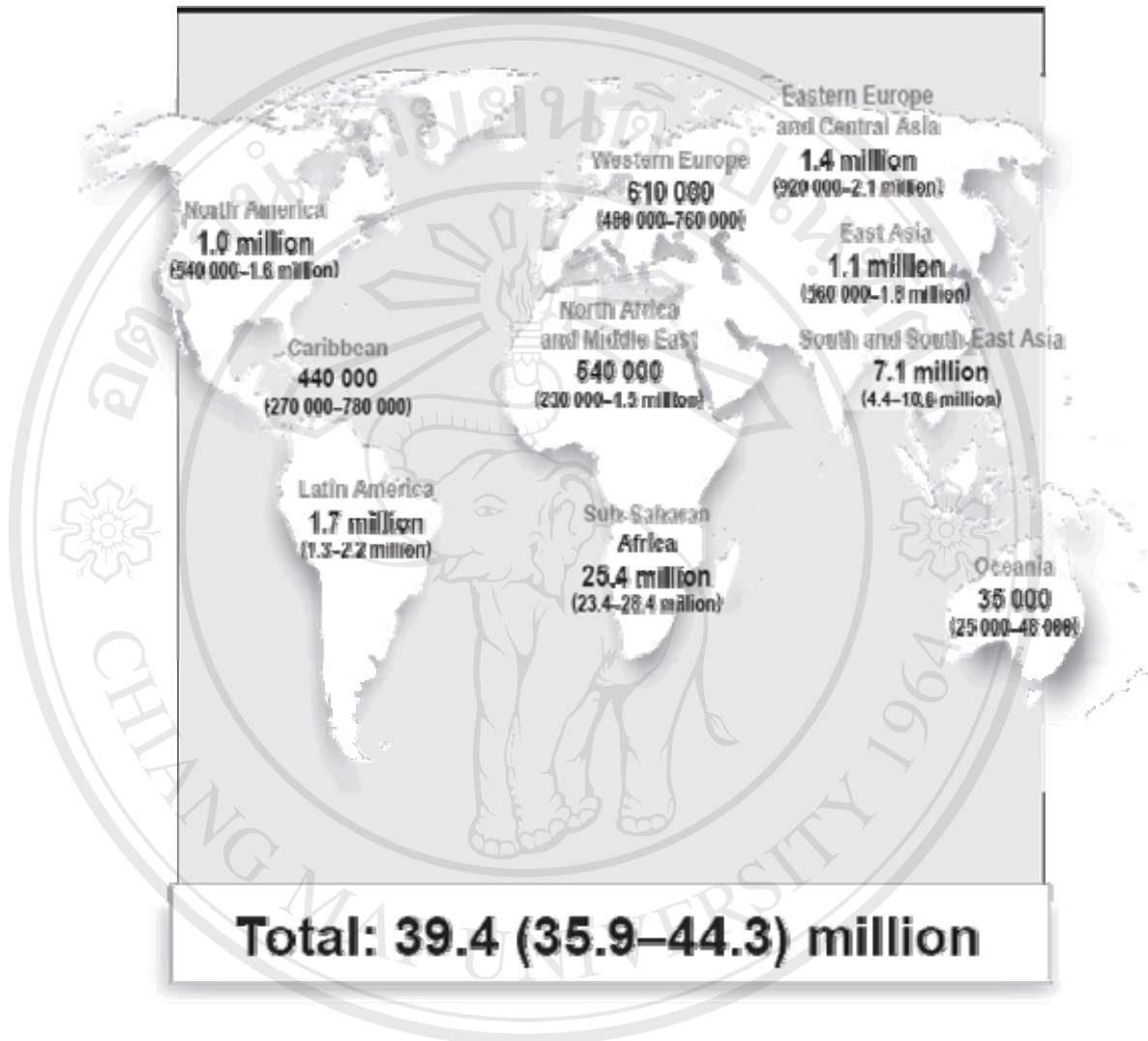


Figure 2.7 Adults and children estimated to be living with HIV/AIDS, end 2004

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
 Copyright © by Chiang Mai University
 All rights reserved

(Source: <http://www.unaids.org>)

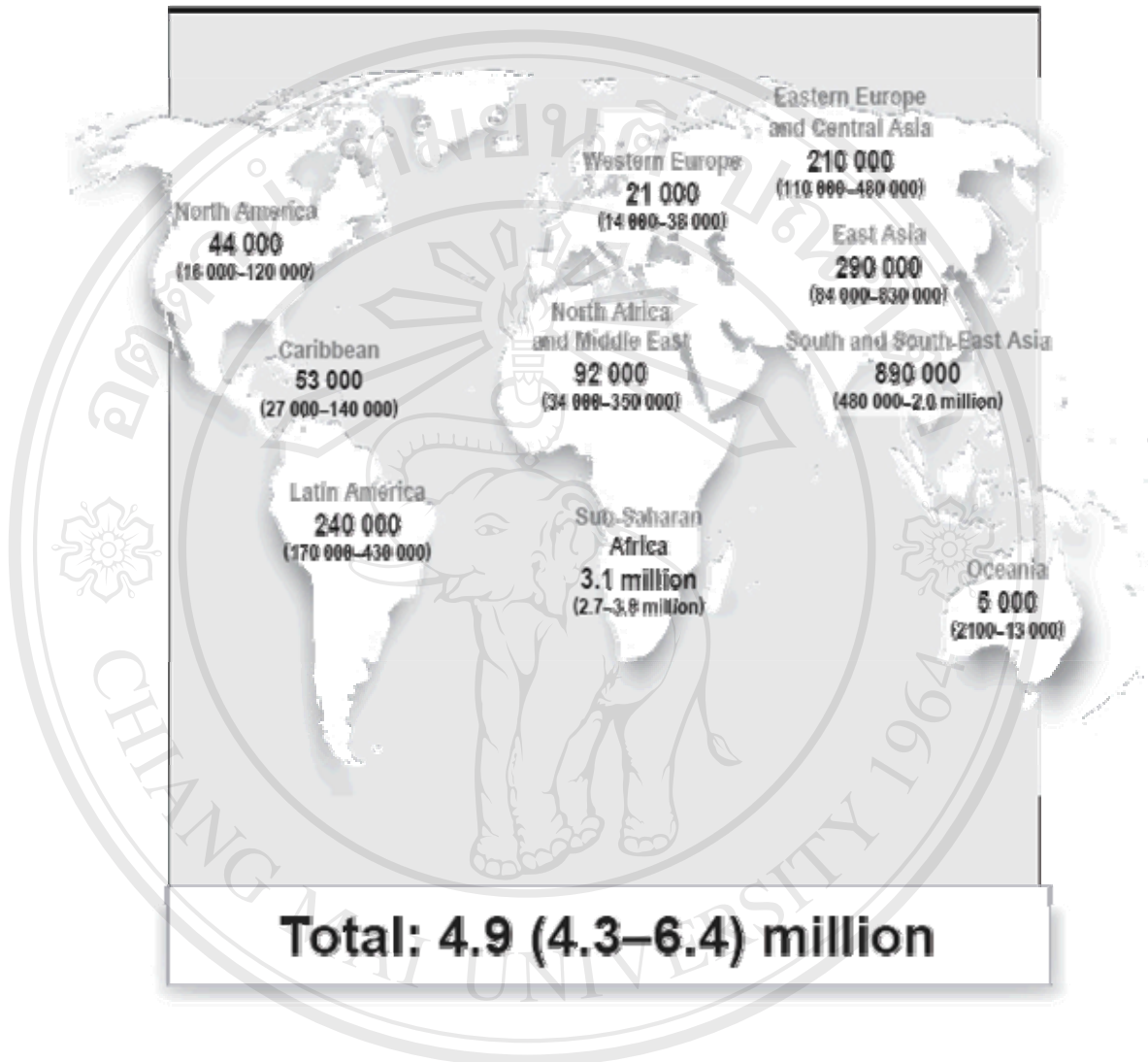


Figure 2.8 Estimated number of adults and children newly infected with HIV during 2004

(Source: <http://www.unaids.org>)

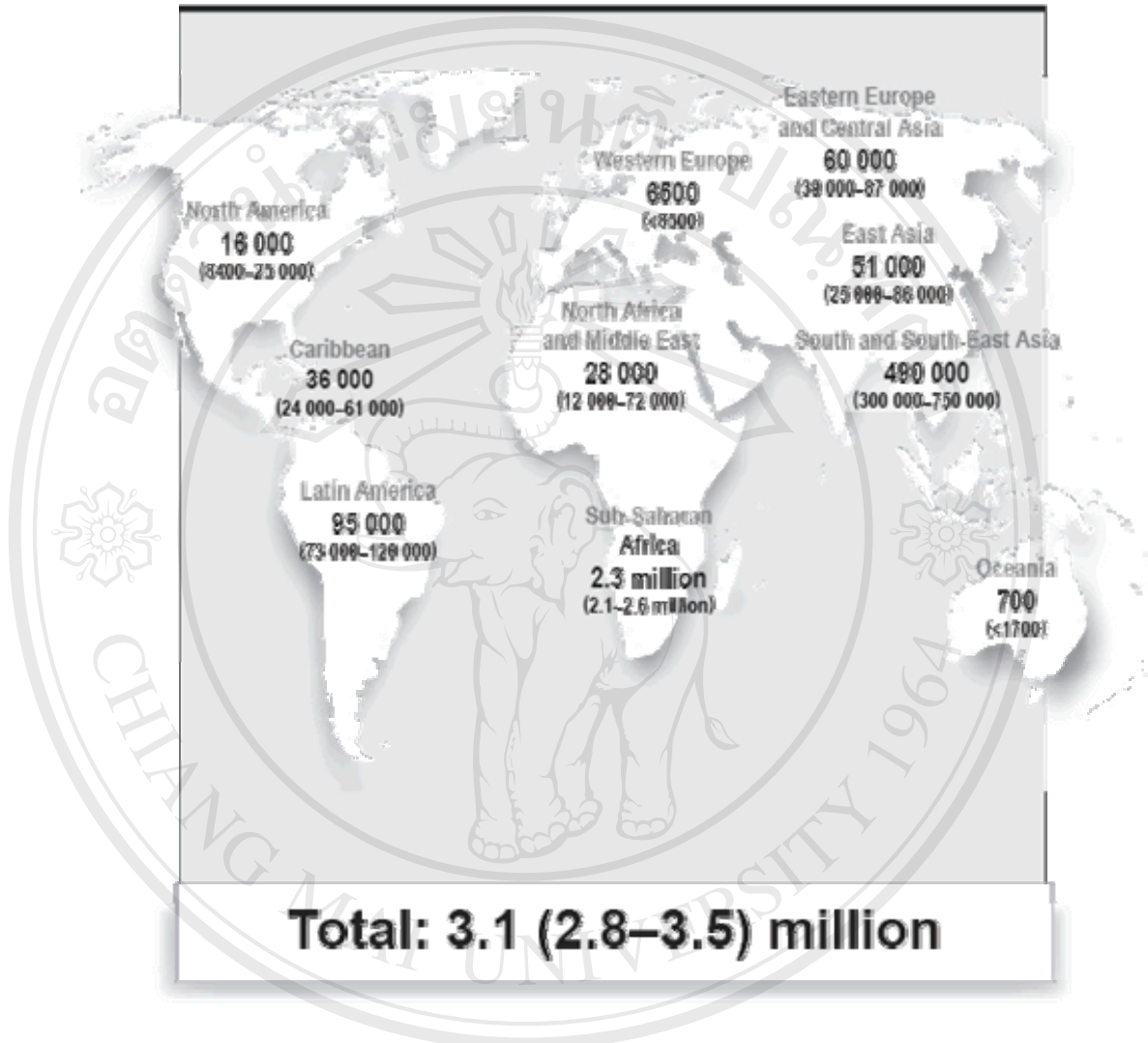


Figure 2.9 Estimated adults and children deaths due to HIV/AIDS during 2004

(Source: <http://www.unaids.org>)

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright © by Chiang Mai University
All rights reserved

2.2.2 HIV-1 subtype and CRF in pandemic

HIV-1 strains are genetically diverse (10, 11, 13, 36), and they can be classified into three groups: M, N, and O. The majority of HIV-1 strains found worldwide, and responsible for the pandemic are group M (Major). Group O (Outlier) seems to be endemic to Cameroon, and neighboring countries in West Central Africa.(56-58) Group N (Non-M, Non-O) has only recently been identified from Cameroonian patients.(59-61)

Within group M, there are nine subtypes (A-D, F-H, J and K), and at least 33 Circulating Recombinant Forms (CRFs).(62) Figure 2.10 shows a phylogenetic tree of HIV-1 strains in group M. These subtypes represent different lineage of HIV, and have some geographical associations. Figure 2.11 shows a global distribution of subtypes and recombinant strains. Subtype B is found mainly in Western Europe, the Americas, and the Australia, whereas subtype A, C and D predominate in East and South Africa. (63) Subtype C causes a wide spread epidemic in India while subtype B, C, CRF01_AE (commonly referred to as subtype E, but actually a CRF of subtype A and E), and their recombinant are found commonly in eastern and southeastern Asia. (63)

HIV-1 subtypes, sub-subtypes, CRF

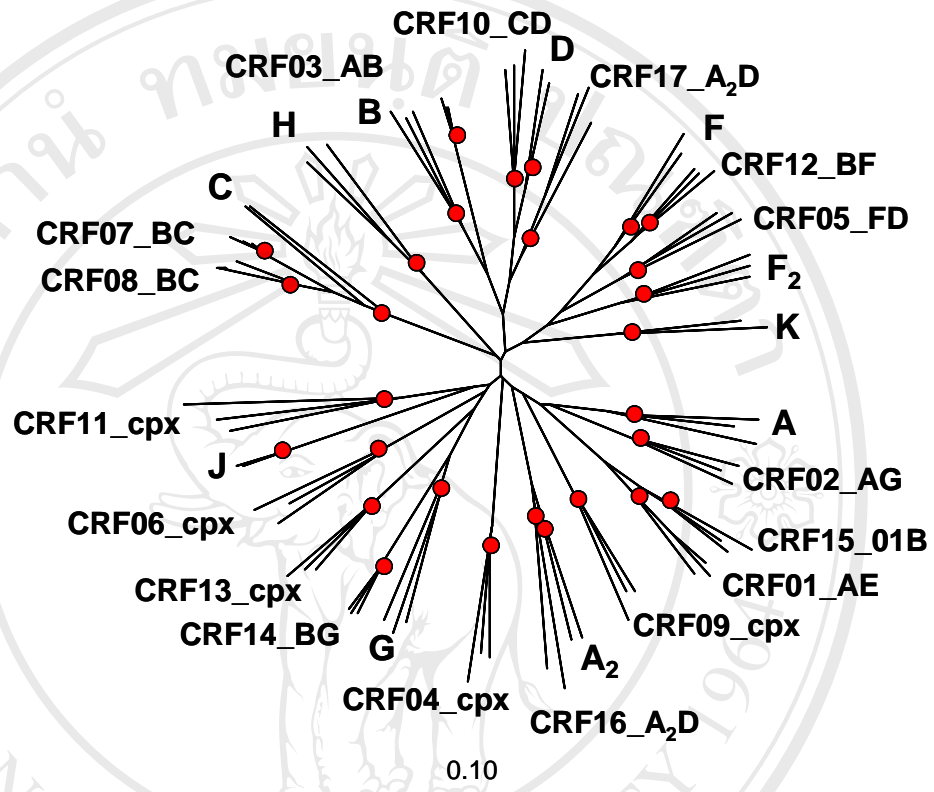


Figure 2.10 Phylogenetic tree of HIV-1 group M

(Source: Francine E. McCutchan, Henry M. Jackson Foundation, US Military

HIV Research Program, Rockville, MD USA)

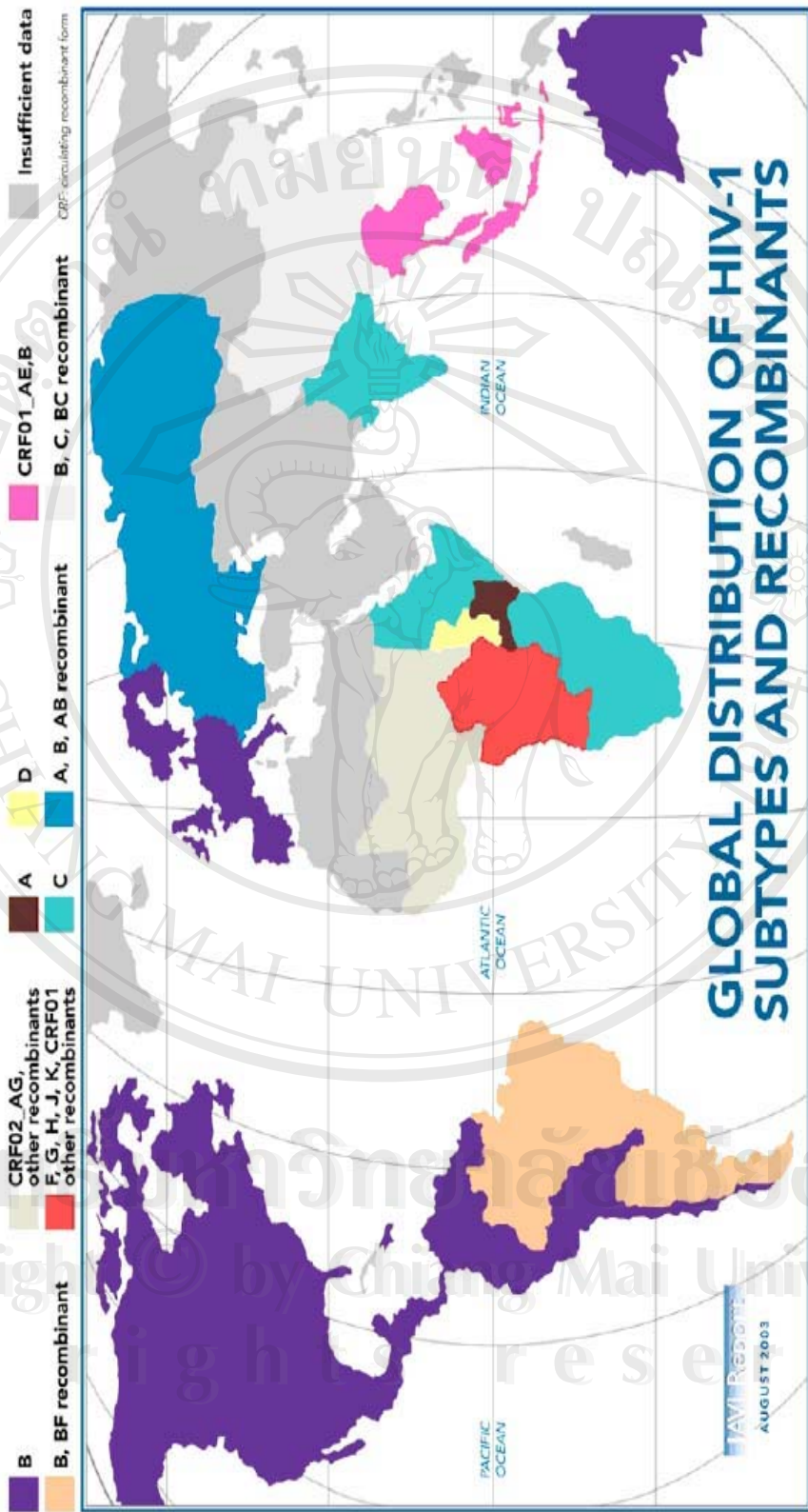


Figure 2.11 Global Distribution of HIV-1 subtypes and Recombinants

(Source : Francine E. McCutchan, Henry M. Jackson Foundation, US Military HIV Research Program, Rockville, MD USA)

2.2.3 The HIV epidemic in South, Southeast Asia and China.

In South and Southeast Asia, AIDS epidemic contains a variety of HIV-1 infections. This area has a higher number of total HIV infections and annual AIDS deaths than any other regions except sub-Saharan Africa. The AIDS picture in South Asia shows dominantly the epidemic in India. From the latest report, there were about 5.1 million (2.5–8.5 million) people living with HIV in India in 2003. (1). The update HIV/AIDS epidemic at the end 2004 in South and Southeast Asia was shown in Table 2.2. The HIV molecular picture in South Asia is principally subtype C alongside subtype B in some regions(65-67), whereas CRF01_AE predominates in Southeast Asia. (64, 68-73)

In Myanmar, the first case of HIV-1 infection was reported in 1989 among injecting drug users (IDU), and later disperse promptly into various risk populations.(74-76) Approximate 68 % of HIV cases are heterosexual transmissions and 30 % to IDU. UNAIDS estimated that 530,000 HIV cases were reported in 1999. (77).The latest report showed that about 330,000 adults and children were living with HIV at the end of year 2003.(64) The HIV epidemic at the end of 2003 was shown in Table 2.3. The highest prevalence of HIV-1 was found in Mandalay in Central Myanmar.(75, 78) Subtypes B', C, and CRF01_AE were found mainly in this country.(78)

Table 2.2 HIV and AIDS estimates in South, Southeast Asia, end 2004

Adult (15-49) HIV prevalence rate	0.6% (range: 0.4%-0.9%)
Adults and children (0-49) living with HIV	7,100,000 (range: 4,400,000-10,600,000)
Women (15-49) living with HIV	2,100,000 (range: 1,300,000-3,100,000)
Adults and children newly infected with HIV in 2004	890,000 (range: 480,000-2,000,000)
Adults and child deaths due to AIDS in 2004	490,000 (range: 300,000-750,000)

(Source: AIDS epidemic update 2004)

Table 2.3 HIV and AIDS estimates in Myanmar, end 2003

Adult (15-49) HIV prevalence rate	1.2% (range: 0.6%-2.2%)
Adults (15-49) living with HIV	320,000 (range: 170,000-610,000)
Adults and children (0-49) living with HIV	330,000 (range: 170,000-620,000)
Women (15-49) living with HIV	97,000 (range: 51,000-180,000)
AIDS deaths (adults and children) in 2003	20,000 (range: 11,000 -35,000)

(Source: 2004 Report on the global AIDS epidemic)

In China, year 2003, following the survey of WHO and UNAIDS, the Chinese government estimated that there were approximately 840,000 HIV infected people in China, including about 80,000 AIDS patients. (1) This number did not include those who had died before the survey. The total populations of Chinese were approximately at around 1,300 million. The HIV epidemic in the end of 2003 was shown in Table 2.4.

The mostly route of HIV transmission was blood-borne in injecting drug users (IDU), followed by recipients of blood transfusions, and heterosexual, respectively. The first sporadic HIV infections were reported in China between 1985 and 1988. The initial subtypes identified were B and B', followed by C and CRF01_AE. The recombination between subtype B' and subtype C, namely CRF07_BC and CRF08_BC have become prevalent in China alongside the pure subtypes. Rarely, subtypes A, D, F, and G have been found. (79) From the study of HIV-1 distribution in provinces of China, the result indicated that subtype B' was widely distributed throughout the country, but each province showed a characteristic picture of HIV-1 strains. (79) In Yunnan province, more than 80% of HIV-1 infections were reported in China through 1996.(80) Various HIV-1 strains have been detected among IDU including subtype B, B'(69), C(81, 82), CRF07_BC, and CRF08_BC.(83) The HIV-1 infections in heterosexuals in Kunming were appeared due to the local epidemic. Subtype B' was predominant before the year of 1997, whereas CRF07_BC, and CRF08_BC became predominant in 2002.(84)

Table 2.4 HIV and AIDS estimates in China, end 2003

Adult (15-49) HIV prevalence rate	0.1% (range: 0.1%-0.2%)
Adults (15-49) living with HIV	830,000 (range: 430,000-1,400,000)
Adults and children (0-49) living with HIV	840,000 (range: 430,000-1,500,000)
Women (15-49) living with HIV	190,000 (range: 95,000-320,000)
AIDS deaths (adults and children in 2003)	44,000 (range: 21,000-75,000)

Source: 2004 Report on the global AIDS epidemic

2.2.4 The HIV epidemic in Thailand

In Thailand, the first case of a Thai patient with AIDS was reported in 1984.

(2) The HIV-1 epidemic started blazingly in 1988 among injecting drug user (IDU) population in Bangkok.(5) More than two decades, the epidemic has changed in many phases. By 2003, it was estimated that 1.5% of the Thai population (adults 15-49) were infected with HIV-1. (1)

In Thailand, in the late 1980s, the major routes of transmission were heterosexual and IDU. During 1988-1995, two strains (CRF01_AE, and B) were separated by routes of transmission. CRF01_AE was a major subtype in sexual behavior, whereas subtype B was predominant in injecting drug users. (5, 85, 86)

The initial classification of the heterosexually infected HIV-1 from Thailand as subtype E was based on analysis of envelope region.(87) Later from complete analysis, these strains were found to be subtype A radiation on the gag gene. (88) Thus, the first full genome sequences from Thailand and from central Africa (49, 50) revealed the AE recombinant, CRF01_AE, the first in a series of circulating recombinant forms to be identified. (13) The predominance of CRF01_AE in heterosexually acquired infections was reported in military cohorts (89) fishermen (90), female (91) and male commercial sex workers (92) and other studies of individuals at heterosexual risk.(93, 94)

The principal early strain found in IDU was subtype B. (93 , 95-97) Subtype B strains found in north America and Europe had their typical GPGR motif at the V3 loop of gp 120 envelope protein, whereas some subtype B in Thailand were divergent and had a GPGQ in V3 (86, 98) resulting in being termed as B-prime (B').(99)

After a decade the epidemic in Thailand has changed. Subtype B, initially the predominant strain among IDUs, was co-circulating along with CRF01_AE in this risk group. CRF01_AE accounted for an increasing fraction of new infections, from 2.6% in 1988-1989 to 25.6% in 1990-1991 and 43.8% in 1992-1993. (100) In the period 1995-1998, almost 80% of new infections in IDUs were CRF01_AE (69), and in the north, 90% of prevalent infections in IDU were also CRF01_AE.(68)

The intermixing of strains led to the emergence of recombinants between CRF01_AE and subtype B. Molecular epidemiology studies revealed the existence of CRF01_AE/B recombinants by 1997, not only in individuals with IDU exposures (101) but, significantly, also in those with only heterosexual exposure. (102) Recently, several subtype-discordant cases detected by screening assays turned out to be the same CRF01_AE/B recombinant form, now called CRF15_01B. (54) Finally, CM 237, one of the very first HIV-1 strains collected from northern Thailand and initially characterized as CRF01_AE by envelope sequencing, has been shown to be one of the unique CRF01_AE/B recombinants, revealing their presence almost from the inception of the Thailand epidemic. (103) The first report of dual infection in Thailand was shown in 1994.(44) Recently, there were reports of two novel CRF01_AE/B recombinant strains, one with the history of intravenous drug use and heterosexual risk behavior (101) and the other with only heterosexual exposure.(102)

The HIV molecular picture in Thailand is changing. CRF01_AE is still predominant, it has recombined substantially with subtype B, and to some extent with subtype C, while the subtype B component of the epidemic may be decreasing. (24, 104) Furthermore, the active trade across the borders of northern Thai province with Myanmar, Laos, and the southern provinces of China, where subtype C, CRF01_AE

and BC recombinants circulate among IDU (14) are likely to facilitate the introduction of new strains and further complicate the epidemic.

2.3 Genetic subtype screening assay

2.3.1 The heteroduplex mobility assay (HMA)

The heteroduplex mobility assay (HMA) was used to investigate the genetic variability of various strains or quasispecies of viruses. The principle of this technique is based on the structural deformations in double stranded DNA, that result from mismatches and nucleotide insertions or deletions caused a reduction of mobility of these fragments in polyacrylamide gel electrophoresis, but not in standard agarose gels. This technique can detect the genetic relationship between multiple viral DNA template molecules. However, it is laborious, limited to one or two small regions of the detected genome, allows the test of only a limited number of samples, and may be difficult to interpret. (17-20)

2.3.2 Subtype specific PCR assay

In addition, subtype specific PCR assay (23 , 105 , 106) which involves PCR and characterization by subtype specific restriction endonuclease sites (PCR-RFLP) (107 , 108), have also been utilized for detecting HIV-1 subtype B or CRF01_AE. The principle of subtype-specific PCR is to amplify the DNA target in some sub-genomic of HIV sequence by HIV subtype-specific primers, while PCR-RFLP base on the

unique patterns of restriction endonuclease site which is generated in each subtype. These are a simple method, but they are generally not sufficient for detection of diverse types of recombinant genomes.(64)

2.3.3 V3-loop peptide enzyme linked immunosorbent assay (V3-PEIA)

HIV-1 serotyping in the name of V3-loop peptide-enzyme linked immunosorbent assay (V3-PEIA) is used to distinguish HIV subtype on the basis of antibody binding to V3 peptides derived from envelop glycoprotein. The sensitivity and specificity of the serotyping assay varied considerably. (22, 23)

2.3.4 The Multi-region Hybridization Assay (MHA)

Recently, a Multi-region Hybridization Assay (MHA) has been developed to determine subtypes, inter-subtype recombinant strains, and dual infections, in East Africa, where subtypes A, C, and D co-circulate.(25) After that, MHA family has been respectively developed to determine subtype B, and F in South America (109), and subtypes A, G, and CRF02_AG in West/West Central Africa, and to distinguish subtype B, C, and CRF01_AE in Thailand, and neighboring countries.(24, 110) The principle of the MHA is to amplify four to eight short regions distributed along the HIV genome in separate first-round PCR, and then each amplicon is amplified the second round PCR in separate tube up to the number of detected subtype, each tube could be detected HIV subtype by using the subtype –specific probe, which was labeled with 6-carboxyfluorescein at 5' end, and had a quencher at 3'end. An

advantage of MHA over other assays is detecting subtype in four to eight regions distributed along the HIV-1 genome. Thus, it should be possible to detect whether inter-subtype recombinant strains and dual infections. Moreover, it is a high-throughput assay which is appropriate in large-scale samples. However, this method requires a real time PCR machine which is expensive and is not suitable in limited-resources laboratory.

2.4 Full-length genome PCR amplification and DNA sequencing

It is a gold standard method for HIV genotyping. This technique uses the nested PCR strategy with near end point dilution of DNA template to recover the amplicon of full-length HIV-1 genome. (27, 48) The first round primers were designed in the 5' and 3' long terminal repeat (LTR), while the second round primers were in the U5 region of 5' LTR and U3 region of 3' LTR. The amplicons obtained from multiple second round PCR were pooled and purified and directly sequenced on both strands. DNA sequences were assembled using Sequencher software version 3.1 (Genecode Inc., Ann Arbor, MI). However, DNA sequencing needs more specific skills, sophisticated facilities and procedure. Therefore, it is not workable as an ordinary tool for HIV genotyping in developing countries.