CHAPTER 2

Literature Reviews

2.1 History of HIV and AIDS

Human immunodeficiency virus (HIV) is an enveloped retrovirus in the genus *Lentivirus,* from the Latin word *lentus* meaning slow, mainly because the virus is persistent and continues to replicate for many years before the progression of disease. HIV is a member of the family *Retroviridae,* and has the reverse transcriptase (RT) enzyme, which converses of ribonucleic acid (RNA) to deoxyribonucleic acid (DNA) in the viral replication. HIV infects host cells, which express CD4, such as CD4+T cells, dendritic cells, and macrophages [23]. In long-term infected patients, the number of CD4+ T cells decline to the point (~200 cells/μL) and the patient becomes susceptible to opportunistic infections, which is the primary characterization of the acquired immune deficiency syndrome (AIDS) stage.

AIDS was first reported in homosexual men who were *Pneumocystis carinii* pneumonia (PCP) and Kaposi's sarcoma in 1981 by Centers for Disease Control and Prevention (CDC) of the United States [24; 25]. Those cases suggested the existence of a new infectious disease based on the patients'resistance to any treatment, the rapid decrease in the level of CD4+ T cells, as well as the presence of opportunistic infections and various malignancies, which associated with AIDS progression, but were not identified as such at that time. This disease was then referred to as Lymphadenopathy because of the characteristic of symptoms.

Luc Montagnier and his colleagues at the Institute Pasteur, France in 1983, first identified the HIV virus. The virus was isolated from a lymph node biopsy of a young homosexual man with a lymphadenopathy syndrome, and was then named as Lymphadenopathy-associated virus (LAV) [26]. In 1984, Robert Gallo and his colleagues at the National Institutes of Health (NIH), USA [27], identified Human T-lymphotropic virus type III (HTLV-III) that actually caused AIDS. In addition,

Jay Levy and his colleagues at the University of California found the AIDS-associated retrovirus (ARV) [28]. In 1986, the International Committee on the Taxonomy of Viruses concluded the viruses which were then named LAV, HTLV-III and ARV, were the same virus that is now recognized as the Human immunodeficiency virus (HIV) [29].

In 1985, HIV-2 was found in Senegalese sex workers that showed cross-reactivity to the simian immunodeficiency virus (SIV) antigens [30]. The virus was isolated and re-characterized in West African AIDS patients, and then referred to as Lymphadenopathy-associated virus LAV type 2 [31], which LAV-2 is now called HIV-2.

The discovery that HIV caused AIDS resulted in Luc Montagnier and Françoise Barré-Sinoussi awarded the Nobel Prize in Physiology or Medicine in 2008 for their research twenty-five years prior.

2.2 Human Immunodeficiency Virus (HIV)

2.2.1 HIV particle

HIV virion has spherical in shape with a diameter of about 100-130 nm. The HIV particle is surrounded by the outer envelope glycoprotein (Env, gp160) that is a bilayer phospholipids obtained from the cell membrane of the host cell. This enveloped protein consists of three surface glycoprotein (SU, gp120) molecules, and three transmembrane glycoprotein (TM, gp41) molecules that anchor via non-covalently links into the viral envelope [32]. HIV enters host cells via interaction between the gp120 with the CD4 receptor and chemokine receptors.

The matrix protein (MA, p17) is located under the outer envelope glycoprotein as it ensures the integrity of the HIV particle. The matrix protein surrounds the capsid protein and proteases enzyme (PR, p10). The cone-shaped capsid protein (CA, p24) is the center of the virion. The capsid also contains two copies of positive single-stranded RNA (ssRNA), nucleocapsid protein (NC, p7) and essential enzymes, which are required for the initial steps of infection and genome replication such as

reverse transcriptase (RT, p66/p51), and integrase (IN, p32). The accessory proteins consist of Nef, Vif, and Vpr that are packed within the virion. The structure of HIV-1 particle is shown in figure 2.1.

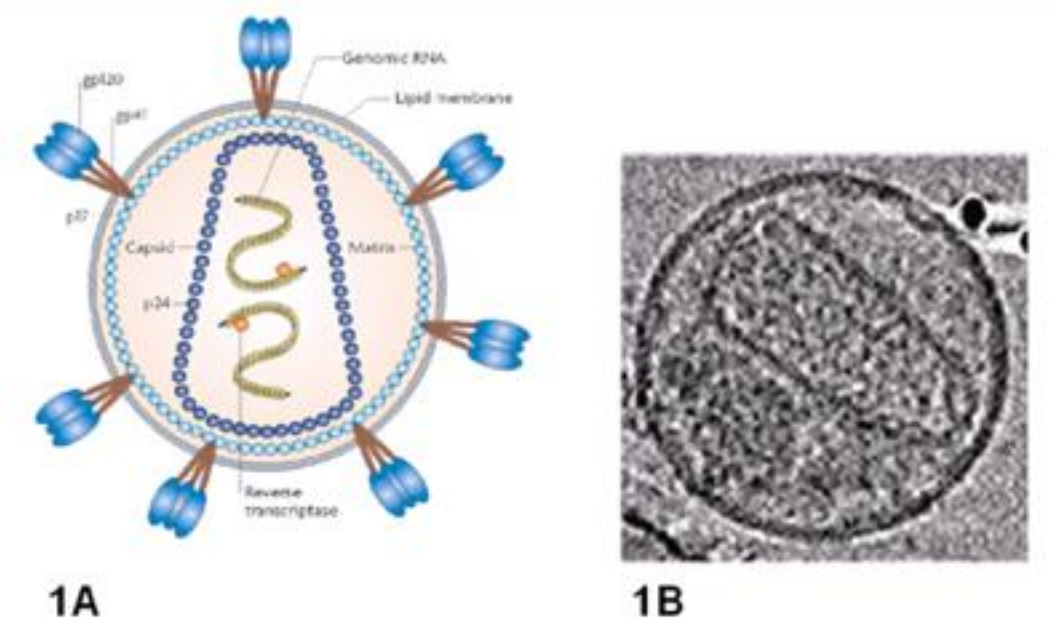

Figure 2.1 The structure of a mature HIV-1 particle is shown. (1A) each virion of HIV-1 contains two copies of an RNA genome and numerous copies of essential enzymes, reverse transcriptase, integrase, and viral protease enzymes. The enveloped protein consists of three gp120 molecules and three gp41 molecules that form as an HIV-1 spike [33]. (1B) electron cryotomography of mature HIV-1 particle [34].

2.2.2 HIV genome

The genome of HIV is made of two copies of a single-stranded positive sense RNA molecule, which have 9,749 nucleotides in length, and encodes the nine genes with flanking long terminal repeat (LTR) sequences at each end of the genome. There are three structural genes *gag*, *pol* and *env* genes, which contain information needed to make the structural proteins for new HIV virion. The *gag* gene encodes for the capsid proteins (group specific antigens). The gag precursor (p55) is a protein that is processed to matrix (MA, p17), capsid (CA, p24), and nucleocapsid (NC, p7).

The *pol* gene encodes for essential enzymes that are needed for the development of the virion such as reverse transcriptase (RT, p66/p51), protease (PR, p10) and integrase (IN, p32) [35].

The *env* gene encodes for viral envelope glycoproteins called gp160, that are cleaved by protease to form non-covalent complexes of the surface glycoproteins gp120 and the transmembrane glycoproteins gp 41 as a trimer that is called a spike at the virion surface [32]. Because the envelope interact with cellular receptors and initial the fusion between the viral and cell membranes, so that are important target of drugs for blocking the early step of the viral replicate. Whereas HIV-1 has many mechanisms for protect them from antibody binding, for example; to generates variants nucleotide substitutions, insertions and deletions of *env* gene, and recombination. Moreover, HIV-1 env spikes are the only obtainable target for neutralizing antibodies [36]. Thus, there are many studies of genetic characteristics of HIV-1 envelope to create the new drugs and develop vaccine that can induce the broadly neutralizing antibodies [36; 37].

In addition, the six accessory genes flank and overlap the *env* gene. They function to regulate the replication of HIV, to inactivate host antiviral functions, and to make HIV-1 replication more efficient. The two regulatory genes contain *tat* and *rev*. Tat enhances the transcription of viral RNA by binding to the transcriptional activation region (TAR) in the provirus 5'LTR to phosphorylates RNA polymerase, and stimulates its RNA elongation activity. Rev binds to the rev response element (RRE) that allows the export of unspliced and partially spliced transcripts RNA from the nucleus that can be translated to produce the structural component of HIV.

The four accessory genes contain *vif*, *vpr*,*nef* and *vpu* (or *vpx* in the case of HIV-2), for which the completion of the virus replication also relies. Vif acts to suppress natural cellular defense against retroviruses. Human cells express a cytidine deaminase called APOBEC that catalyzes the conversion of deoxycytidine to deoxyuridine in the first strand of viral cDNA, which eliminates its efficacy to encode viral proteins. Vif leads the transport of APOBEC to proteasomes where it is divested [38; 39]. Vpr has various functions that enhance viral production and release, for example the induction of cell cycle arrest and nuclear import of preintegration

complexes. Vpu is found in HIV-1, but not in HIV-2, and represents this accessory gene as Vpx. Vpu is unique to HIV-1 and variants of SIV, which is desired for the maturation of progeny virions and the enhancement of virion release [40]. Also, Vpu degrades CD4 in the endoplasmic reticulum [40; 41]. Nef induces T-cell activation and the persistence state of HIV infection [42; 43]. The attributes of the HIV-1 genome are shown in figure 2.2.

Figure 2.2 Illustrates the HIV-1 genome [44] that is flanked by long terminal repeats (LTR). The genome can be read in three frames, and several of the viral genes overlap in different reading frames. HIV-1 has three major genes- *gag*,*pol*, and *env*. HIV-1 has six smaller genes that encode proteins, which affect viral replication and infection. Two of these, *tat* and *rev*, perform regulatory functions. The other four accessory genes contain *nef*, *vif*, *vpr* and *vpu*.

2.2.3 HIV replication

HIV enters the CD4+ cells of the immune system by binding of the viral envelope glycoprotein gp120 to CD4 with high affinity. The gp120 composes of five constant (C1-C5) and five variable regions (V1-V5). This results in conformational changes in gp120 that expose the binding sites (V3) for chemokine receptors or co-receptors of HIV-1 that rely on the viral tropism, which are usually CCR5 or CXCR4. This second binding event induces a second conformational change in gp120 that releases the transmembrane glycoprotein gp41 to bind the cellular membrane of the host cell, which causes fusion of the virus membrane with the cellular membrane to allow viral genome and the enzymes to enter the cell [45].

Once in the cytoplasm [46], reverse transcription of the viral RNA genome is conducted by reverse transcriptase [47]. This enzyme has two enzymatic activities composed of DNA polymerase, which copy RNA or DNA template into

a complementary DNA sequence. In addition, RNase H acts to degrade the RNA strand of an RNA-DNA duplex into small pieces as template for the first DNA strand [48]. These results in the production of double-stranded DNA (dsDNA) and formation of the pre-integration complex (PIC) contain two copies of the viral DNA genome, the matrix protein (MA), Vpr, and the enzymes. PIC is translocated from the cytoplasm to the nucleus.

In this step, Vif acts to suppress natural cellular defenses against retroviruses by leading APOBEC to proteasomes [38; 49]. Viral dsDNA enters the nucleus by Vpr activity and is inserted into the host genome as the HIV provirus, via the integrase enzyme.

Activation of CD4 T cells leads to the expression of the transcription factors that bind to the proviral LTR and initiate transcription by RNA polymerase II, and the production of viral mRNAs. Tat bind to the transcriptional activation region (TAR) in the provirus 5'LTR to enhance the transcription of viral RNA by two ways. One way is phosphorylate RNA polymerase, and the other is to stimulate its RNA elongation activity [50]. Rev acts to export some of mRNAs from the nucleus into the cytoplasm that are translated to produce the structural proteins, core, and envelope of the virus. The new viral genomes are packaged with these proteins to form many new virus particles, and the immature HIV viruses are assembled and bud from the host cell. Internal assembly then occurs with the cleavage of a large capsid into the small capsid by the protease enzyme, producing mature virus particles that are released and can infect additional cells. The HIV-1 replication cycle is shown in figure 2.3.

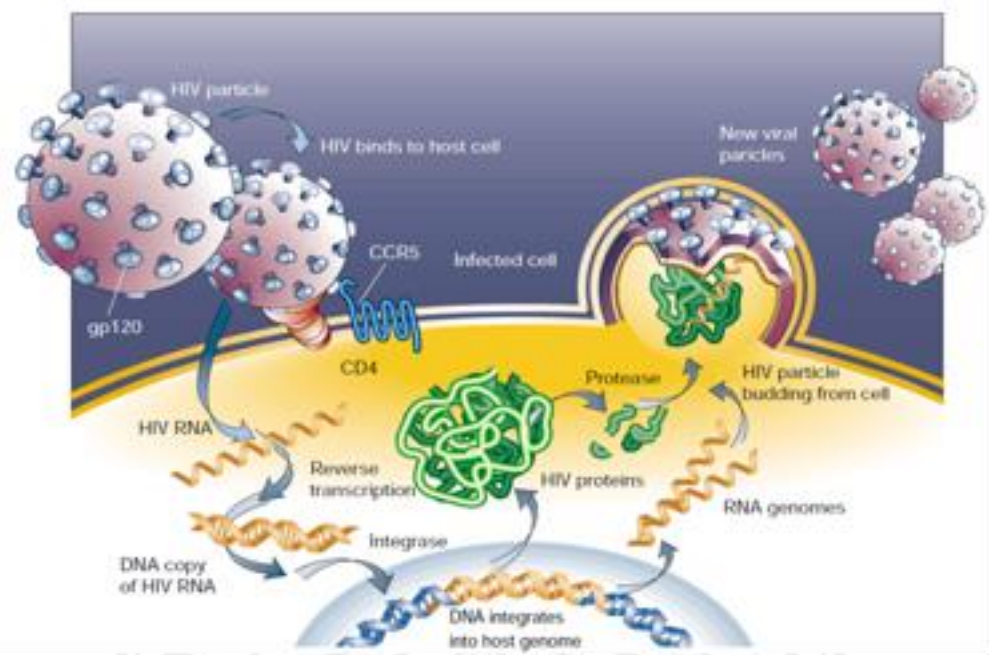

Figure 2.3 Demonstrates the HIV-1 life cycle in the host cell. After binding with CD4 and the co-receptor, HIV genome fuses into the host cell. Reverse transcription of the viral RNA genome result in the productions of double-stranded DNA (dsDNA), which are integrated into the host genome. Activated CD4 leads to the transcription and the production of viral mRNAs. The new viral genomes are packaged with these proteins to form new virus particles. The immature HIV viruses are assembled, bud from the host cell, and produced mature virus particles by protease enzyme [45].

2.3 Genetic diversity of HIV

2.3.1 HIV classifications

HIV maintains a genetic diversity that leads to the diversification of the HIV population. HIV is a member of the family *Retroviridae*, genus *Lentivirus*. HIV is classified on the basis of serologic properties and nucleotide sequences, into two types, HIV-1 and HIV-2, have genetic homology of about 40% to 50%, with the greatest variable sequence in the *env* gene. Both viruses seem to have spread to humans from other primate species.

HIV-1 has passed to humans from chimpanzees (SIVcpz) [51; 52], while HIV-2 originated in the sooty mangabeys (SIVsm) [53; 54]. Most AIDS worldwide is caused by the more virulent HIV-1, whereas HIV-2 is endemic in West Africa and is now being spread around India. The different HIV-1 sequences are classified into different groups, subtypes, sub-subtypes, and circulating recombinant forms (CRFs) or unique recombinant forms (URFs). HIV-1 can be classified into four groups: M (major), O (outlier), N (new, or non-M, non-O) and the recently identified P [55; 56]. The M group is the major cause of AIDS worldwide [57], whereas groups O, N and P are less common. Group O has been endemic in West Central Africa and Cameroon [58; 59]. Group N has been found in Cameroonian patients [55], and group P was recently identified in two patients from Cameroon as well [55; 56].

The pandemic HIV-1 group M consists of nine subtypes (A to D, F to H, J, and K). Subtypes A and F are also subdivided into sub-subtypes, including sub-subtypes A1 to A5 [60; 61], and sub-subtypes F1, F2. Within a subtype, variations at the amino acid level occurs in about 8-17%, but can be as high as 30%. Whereas variations between subtypes are usually between 15-20% for the *gag* and *pol* gene, and 20-30% for the *env* gene [62, 63], but can also be as high as 42%, depending on the subtypes and genome regions examined [64]. Subtypes within groups O and N are not yet identified. A phylogenetic tree representing the HIV genetic diversity is shown in Figure 2.4.

Figure 2.4 Phylogenetic relationships of HIV groups M, N, O, CRFs, and the distance between HIV-1 and HIV-2 [65]

2.3.2 HIV recombinants

Advances in full-genome sequencing of HIV have led to the identification of the intersubtype recombinant genomes becoming designated as circulating recombinant forms (CRFs) or unique recombinant forms (URFs). CRFs or URFs are mosaic strains with segments from two or more subtypes alternating across the genome, which is a significant mechanism to evade host immunity [66] or antiretroviral therapy [67]. CRFs are identified in three or more people with no direct linkage and characterized as the same recombinant structure by full-length genome sequencing. Whereas URFs are recovered from only a single or two people [68]. The CRF's assigned name reflects the sequence of discovery, and reported in the literature and subtype composition, starting with CRF01_AE. If the recombinants consist of contributions from three or more different subtypes, then they are replaced by designation "cpx", denoting complex, e.g. CRF04_cpx (A, G, H, K, and U). Currently, 72 CRFs (CRF01 to CRF72) have been engaged in the HIV database at Los Alamos National Laboratory (www.hiv.lanl.gov).

2.3.3 Mechanisms of genetic diversity

The high genetic diversity in the form of viral quasispecies is one of the causes of HIV's spread worldwide [69]. HIV genetic variability depends on high mutation [70] and retroviral recombination rates of the reverse transcriptase enzyme [47], as well as a high replication rate [1]. Whereas the selection from the host immune system [37], and genetic and phenotypic constrains to variation are also the factors of hallmarks in HIV [71]. These mechanisms result in genetically diverse viral populations in each infected individual, term "quasispecies". The viral sequences within a single individual can differ by up to 10% [72]. The genetic diversity is associated with viral persistence in the immune selection, the failure of antiretroviral therapy, and the problem in vaccine development.

2.3.3.1 High mutation rate of reverse transcriptase

The activity of the reverse transcriptase (RT) for viral replication is the conversion of genome RNA into dsDNA by RNA-dependent and DNA-dependent DNA polymerases activity, and RNase H activity. However, the RT lacks a proofreading activity, 3' exonuclease activity, and the ability to confirm an accurate copy of the RNA code. RT can introduce nucleotide substitutions approximately 10^{-4} per nucleotide per replication cycle. The mutation rate of HIV-1 is about 3.4×10^{-5} mutations per base pair per replication cycle [70], or introduce on average 1 mutation (range, 0 to 3) for each viral genome transcribed [73]. Currently, the mutation rate of the HIV-1 genome can be as high as 5.4×10^{-5} mutations per base pair per replication cycle by complete sequences analyzing or near-full-length genomes [74].

2.3.3.2 High turnover rate of HIV

The high replication rate of HIV contributes to generating viral quasispecies. HIV-1 has a rapid turnover rate at about 10.3×10^{9} virions per day. The average lifespan of virus producing cells is 2.2 days (half-life approximately 1.6 days), whereas plasma virions have a life-span of 0.3 days (half-life approximately 0.24 days) [1; 2]. The wild-type strains are replaced completely by drug resistant virus occurring in plasma within 2-4 weeks [46]. In one day, millions of HIV variants are created within

any infected person.

2.3.3.3 The generation of HIV recombinant

HIV-1 recombination can lead to viral diversity that occurs when one person has dual infections with two or more HIV strains derived from two different individuals, which may be the same subtype (called intrasubtype recombination) or different subtypes (called intersubtype recombination). Dual infections can be a result of simultaneous infections with two different strains (coinfection) before seroconversion. Whereas super-infections are due to sequential infections, in which a second infection with a heterologous strain comes after seroconversion of the initial infection. Dual infection may be a result of a higher viral load and a more rapid disease progression [75, 76].

The HIV-1 life cycle with co-infection of two HIV strains is shown in figure 2.5. When the host cell is infected with two different strains (subtype X and Y), the RNA genomes of two subtypes are reverse transcribed into dsDNA. Proviral DNA segments from X and Y are then integrated into the host genome. After the transcription and the production of viral mRNAs and proteins, X and Y RNA strands are packaged into the same virion that is called heterozygous particle. Then this heterozygous particle infects the second host cell, X and Y RNA are then reverse transcribed into mosaic DNA by error-proof reading RT, which results recombination. XY mosaic DNA is integrated into the host genome, and is transcribed to RNA and proteins. Finally XY RNA and proteins are assembled and bud to form the recombinant virus.

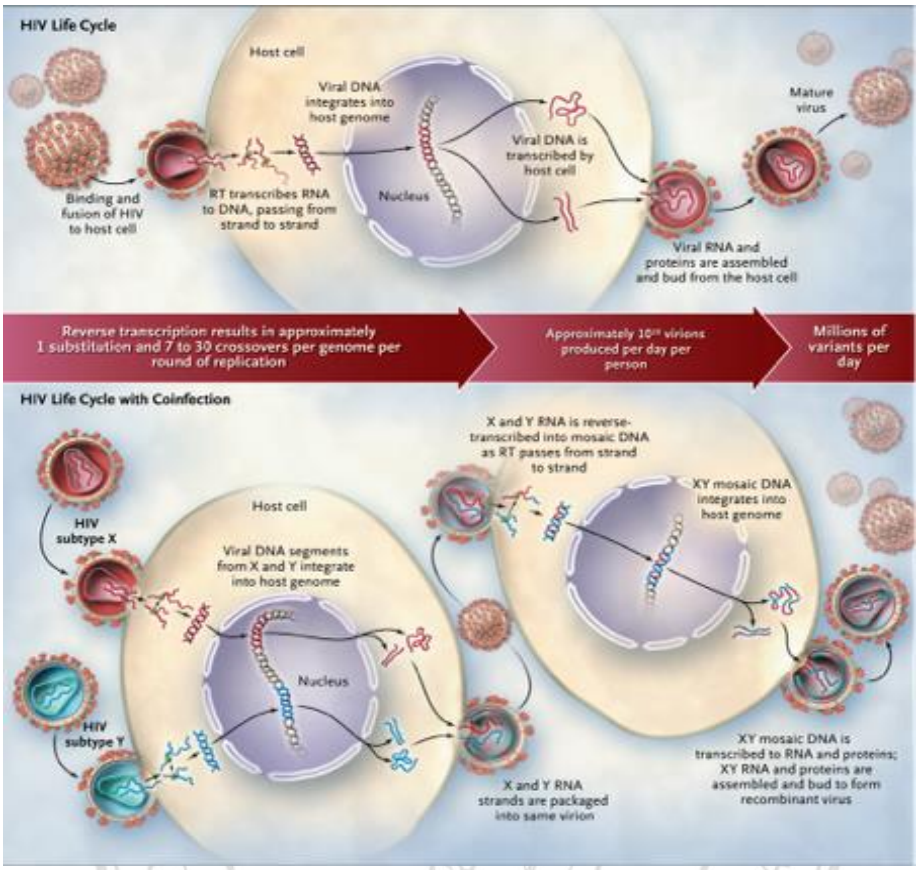

Figure 2.5 The generation of HIV recombinant [3]. After the host cell is infected with two different strains (subtype X and Y), proviral DNA segments from X and Y are integrated into the host genome. Then X and Y RNA strands are packaged into the same virion that is called heterozygous particle. When this particle infects the second host cell, the X and Y RNA are reverse transcribed into mosaic DNA by RT, which results in recombination. XY mosaic DNA is integrated into the host genome and is transcribed to RNA and proteins. Finally XY RNA and proteins are assembled, and bud to form a recombinant virus.

At the molecular level, HIV-1 recombinant process can occur in two different stages; during the synthesis of the minus strand DNA or plus strand DNA. The placement of multiple selection factors is a result, which are three major processes. The first factor includes the reverse transcription step that generates recombinant genomes by the copy choice mechanism. The second factor is the selection for the functional forms. 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 [77]. The new recombinants can be identified by breakpoints that separate different genomic subtype [78; 79]. The sites of recombination are getting randomly along the viral genome. Intersubtype recombination is the most frequently observed, whereas intra-subtype recombination is possible. The HIV-1 recombination rate was about two to three recombination events per genome per replication cycle, and crossovers or breakpoints were identified across HIV-1 genome [80].

A study of analyze the position of the breakpoints in sequences of the Los Alamos Database (http://www.hiv.lanl.gov/) found that recombination breakpoints are non-randomly distributed across the genomes of HIV-1 [81]. The borders of the *env* gene around the first exon of *tat*, *vpu* and the beginning of *env*, and the second exon of *tat*, *rev*, and the 3' end of *env* were identified two recombination prone regions or "hot spots". Moreover, comparing the recombination breakpoints across gp120, the C2 region is a hotspot for recombination [82].

2.4 The implication of HIV epidemiology study

2.4.1 Transmission and disease progression

Different viral subtypes may effect to HIV transmission and disease progression. Many reports show the partition of subtypes to different risk groups. For example, subtype B was more common among IDUs, and CRF01_AE was also more common among heterosexuals in Thailand [83]. In South Africa, subtype B was more associated with homosexuals, and subtype C with heterosexuals [84]. In Argentina, MSM were usually infected with subtype B, while heterosexuals and IDUs, were sustained by BF recombinant [85]. Also, previous studies regarding mother-to-child transmission suggested that subtypes A, C and recombinants were more parentally transmitted than subtype D [86]. Although pregnant women infected with subtype C were more likely to transmit HIV to their children than those infected with subtype B [87]. New CRFs have often been found among IDU, for example CRF03_AB, CRF07_BC, CRF08_BC, CRF14_BG, and CRF35_AD [88, 89].

Several studies reported that the different HIV-1 subtypes may relate to viral load [90; 91], transcriptional activation levels [92], and coreceptor usage [93]. The HIV strains that use CCR5 or R5 viruses are more frequently transmitted than strains that use CXCR4 (X4 viruses). However, X4 viruses emerge later in infected patients and are associated with more rapid disease progression [94]. All HIV-1 subtypes can use both coreceptors, but subtype D may use dualtropic (an R5X4 virus) most frequently, whereas subtype A mostly uses CCR5 in late infection [89; 95]. Moreover, a less common emergence of CXCR4-using (X4) variants was found in HIV-1 subtype A infection, compared with HIV-1 subtype D that use CXCR4 more frequently during early infection [95]. In advanced AIDS patients, X4 viruses in subtype C infected patients were less emerged when compared with subtype B infected patients [96]. The subtype D-infected patients had more rapid progression than those infected with subtype A as were reported in Uganda, Kenya, and Tanzania [97; 98]. Several studies on the HIV-1 subtype differences in disease progression have also been reported. For example, subtypes C and D result in more violent diseases, followed by G, AE, AG and A, result in the least violent of all HIV-1 subtypes [97-100]. A study in Senegal reported that non-A subtype infected women were more susceptible to AIDS than subtype A infection [99]. In addition, a European study showed subtype D had a lower CD4 count compared with other subtypes (A, B, C, and CRF02 AG, which had similar rates of CD4 loss) [101]. In Chinese HIV infected patients, CRF01_AE was associated with faster disease progression when compared with other subtypes [22].

To explain the high virulence of recombinant HIV-1, the report from Uganda showed that within a same population with multiple subtype infection resulted in the highest progression in CD4 cell count of 250 cells/mm^3 or death due to AIDS, following with recombinant forms, HIV-1 subtype D, and subtype A [98]. However, there were the reports showing no differences in disease progression between patients infected with subtypes B and C in Israel [102], patients infected with subtypes A, B, C and D in Sweden [103], subtypes B and CRF01_AE in Thailand [104] and subtypes CRF02_AG and other subtypes in Cameroon [105].

2.4.2 Diagnosis and monitoring tests

There is a high variation of HIV impact on sensitivity and specificity of diagnostic tests, because it detects all variants of HIV-1 and HIV-2. Serologic assays are the screening techniques for diagnostic and blood screening tests. The 1st generation tests detect antibodies to capsid protein p24 *gag* and gp41 *env*, inactivated viruses. The 2nd and 3rd generation tests detect antibodies of synthetic peptides or recombinant protein antigens. HIV-2 and HIV-1 group O were not completely detected in early diagnostic tests by some serologic tests because of gp41 variation [106-108]. Currently, the 4th generation HIV immunoassays (IAs) are designed to detect both the p24 antigen and antibodies in a single test [109]. Whereas antigens can be detected before the formation of antibodies that can reduce the seroconversion window period in newly infected people. Although 4thgeneration IAs are advanced methodologies that are avalable for detecting HIV-1 group M, group O and HIV-2, this assay is not available in some countries, where most new infections originate. For instance, in Africa with a high genetic diversity of HIV, the operation of diagnostic tests is much less sufficient [110]. The study to compare the efficiency of the 2nd, 3rd and 4th generation HIV IAs showed the different sensitivity and specificity for detection of HIV-1 group M, CRF01_AE, CRF02_AG and group O [111].

The study in Cameroon suggested the sensitivity and specificity of diagnostic tests include both five rapid tests that compose of Retrocheck HIV (Qualpro Diagnostics, India), SD Bioline HIV 1/2 3.0 (Standard Diagnostics, South Korea), HIV (1+2) Rapid Test Strip (Shanghai Kehua Bio-engineering, China), Determine HIV-1/2 (Inverness Medical Innovations, USA) and ImmunoComb II HIV1&2 BiSpot (Orgenics, Israel) and two 4th generation IAs, Enzygnost HIV Integral II (Dade Behring, Germany) and Murex HIV Ag/Ab Combination (Murex Biotech, UK). The samples including HIV-1 group M, O and recombinant HIV plasmas were studied. The results showed that in the recombinant HIV plasmas, the specificity of all diagnostic tests varied from 77.9% to 98.0%. Furthermore, two rapid tests, HIV (1+2) Strip and Retrocheck, could not detect all HIV-1 group O samples [110]. Even though the comparative studies showed that most of immunoassays, and especially rapid tests, are sensitive and specific for diagnosing HIV-1 group M infected persons with chronic

stage [107; 112; 113]. They had a lower sensitivity for non-B subtype detection, because these assays and were developed from HIV-1 subtype B strains [114].

Additionally, commercial real time PCR assays for viral load measurements had various results (index of consistency 0.991 to 0.999) to quantify of HIV-1 RNA dilution (316,230 copies/ml to 50 copies/ml) when testing diverse genetic variants of HIV-1 group M and O from Africa, especially group O [115]. The occurrence of new variants that challenge the assays of detection, as non-detection or unsuitable detection, can lead to inaccurate and even low viral load measurements [116]. For instance, the study in Canada reported the H/J recombinant virus was not detecting viral load levels, and also the results of further analysis of the genome sequencing found a mutant gene, specifically the *gag* gene, that was in a specific position with primers and probes in commercial kit, Bio-MerieuxNuclisensHIV-1QT (*BioMerieux*, Netherlands) and Roche Amplicor Monitor HIV-1, v1.5 (Roche Diagnostics, USA) [117].

2.4.3 Response to therapy in HIV-1 subtype

Genetic variability of HIV may impact the response to antiretroviral (ARV) treatment by involving genes that encode viral proteins targeted by ARV drugs. The *pol* gene encodes many enzymes (protease, reverse transcriptase and integrase), and the *env* gene codes for the transmembrane glycoprotein gp41. These HIV-1 proteins are targets for protease inhibitors (PIs), RT inhibitors (nucleoside and non-nucleoside reverse transcriptase inhibitors; NRTIs and NNRTIs), Integrase inhibitors (INIs) and fusion inhibitor (FI), respectively. However, the ARVs are usually developed based on the HIV-1 subtype B, which is 12% of all global infections, and has been widely used in different parts of the world where non-B subtypes are primarily circulating. There were reports that showed different group M subtypes, both B and non-B subtypes that had similar susceptibilities to the use of PIs, NRTIs and NNRTIs [118-120]. Although the combinations of ARV regimens are effective to all HIV-1 group M subtypes, but there is an occurrence of different response of ARV in various subtype infection. For instance, PIs had various efficient therapies in the different subtypes of HIV-1 infection: C, F, G and CRF02_AG [119; 120]. In a comparison of responses to Indinavir, the patients infected with subtype C and G were more responsive than

subtype B. After treatment with Nelfinavir and Ritonavir, CRF02_AG was more susceptible than subtype B, C, F and G as a result of the mutation in position 19, 35, 37, 70 and 89 of *pol* gene encoded protease enzyme [120]. Tipranavir, the protease inhibitor resulted low in efficiency among patients infected with subtype F [119, 120].

HIV develops several mutations for resistance to protease inhibitors due to a high genetic barrier, whereas a single mutation can result in the resistance to NNRTIs because of the low genetic barrier [121]. For example, several studies showed the V106M substitution in subtype C could induce resistance to all NNRTIs [122-124]. While a single transition of the V106M mutation in RT was assisted in subtype C (GTG to ATG); the other subtypes could induce resistant by two transitions (GTA to ATG). Both HIV-1 group O and HIV-2 had a natural resistance to NNRTIs [125; 126].

2.4.4 HIV vaccine

The great barrier in the development of a globally effective HIV vaccine is HIV diversity that is shown by the mutation and circulation of recombinant forms in individuals and populations. After being infected with HIV-1, hosts have cellular and humeral immune responses that are specific of infected strains, but HIV has a mutational escape from response by the CD8+ cytotoxic T cells and neutralizing antibodies [127; 128]. Thus, the idea for controlling the circulating HIV is a preventive vaccine that can stimulate neutralizing antibodies to recognize every subtype of HIV by the response of CD4 T-cells and CD8 T-cells [129; 130].

2.5 Epidemiology of HIV infection and global distribution

The World AIDS Day 2014 Report of UNAIDS highlighted people living with HIV is about 35 million worldwide at the end of 2013. Nearly 2.1 million people were newly infected with HIV, which was 38% lower than in 2001. At the end of 2013, 1.5 million people were dying of AIDS-related a cause, which has fallen by 35% since the peak in 2005 (2.4 million in 2005). There were 24.7 million people living with HIV in sub-Saharan Africa, which accounts for 70.6% of all infection [4].

In Asia and the Pacific, there were 4.8 million people living with HIV at the end of 2013, and an estimated 350,000 new HIV infections, which has declined by 6% between 2005 and 2013. The percentage of people living with HIV by country in ASIA and the Pacific is shown in figure 2.6. While new infections declined in Myanmar, Thailand, Vietnam and India by 58%, 46%, 43%, and 19%, respectively. New HIV infections emerge more than doubled in the Philippines between 2001 and 2012, and in Indonesia have new HIV infections have raised by 48% since 2005. The number of AIDS-related deaths in Asia fell by 37% between 2005 and 2013. Under the 2010 World Health Organization (WHO) guidelines, the overall treatment coverage is approximately 30% in Asia and the Pacific. Only two countries in Asia and the Pacific, Thailand and Cambodia, have more than 50% of all people living with HIV currently on antiretroviral treatment [4].

The Asian Epidemic Model (AEM) Projections for HIV/AIDS in Thailand 2005-2025 and the Thai Working Group on HIV/AIDS Projections estimate the state of the Thai HIV epidemic in 2012, where 464,414 adults (15-49 years) were living with HIV. New infections were about 9,473, and a shift of the dominant modes of HIV transmission from sex work to sex between men, who have sex with men (MSM), followed by sex between married couples, intravenous drug users (IDUs), and then sex with sex work [131].

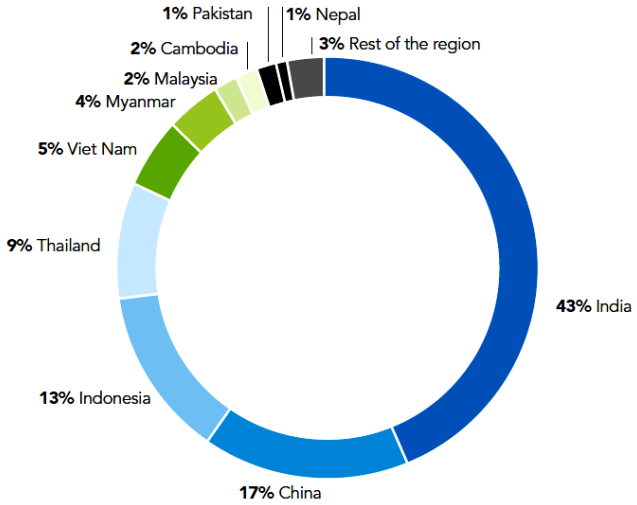

Figure 2.6 The percentage of people living with HIV by country in ASIA and the Pacific. China, India, Indonesia, Myanmar, Thailand, and Vietnam accounting for more than 90% of the people living with HIV in the region. (Based on the subtype distribution available from the Los Alamos HIV Database; http://www.hiv.lanl.gov/).

The involvements of the molecular epidemiology of HIV-1 are an effect of the global distribution of subtypes and circulating recombinant forms. Although HIV-1 genetic subtypes are dispensed in different parts of world, the most widespread HIV-1 subtypes are subtypes A, B and C [72]. Prior to 2005, half of all HIV-1 infections worldwide are subtype C infections that result of the epidemic in Southern Africa, South America and Asia [132-135]. In 2013, the most predominant HIV-1 strain is subtype B, with 56.1% infections of worldwide. The other HIV-1 subtypes for example C, A, D, G, and other subtypes have about 15.9, 7.3, 3.6, 1.1, and 6.3% of total infections, respectively. The distribution of the recombinant forms, three most important CRFs (CRF01_AE, CRF_02AG and CRF07_BC) have prevalence account for 5.7, 2.7, and 1.3% of total HIV-1 infections (based on the subtype distribution available from the Los Alamos HIV Database; http://www.hiv.lanl.gov/). HIV-1 recombination plays a role in HIV epidemics in different regions. Africa, Asia and South America are classified as recombination hot-spots. About 20% of infections worldwide are all recombinant forms [77].

Subtype B is the most widely spread subtype that is most common in North and Latin America, the Caribbean, Europe, and Australia. It is also common in several countries of Southeast Asia, North Africa, the Middle East (Israel), South-Africa and Russia. Subtype C is the strain mainly found in southern Africa, Ethiopia and India. Subtype A viruses are supreme in the central and eastern parts of Africa (Kenya, Rwanda, Uganda, and Tanzania), in eastern European countries, and the Soviet Union. Subtype D viruses are mainly found in East and West Africa. CRF01_AE are the predominant subtypes in South-East Asia, including Thailand. CRF02_AG and other recombinants are the predominant types in West and Central Africa [72].

In South America, a mixture of subtype B and BF recombinants are most prevalent, with a small measure of subtype C infections. In East Asia subtypes B, C and BC are the predominant recombinant strains found. Central Africa found several rare subtypes (F, G, H, J and K) and a variety of recombinants. Unlike other regions of the world, Central Asia has less data about HIV circulation [72].

The global distribution and prevalence of HIV-1 subtypes and recombinant forms are shown in Figure 2.7.

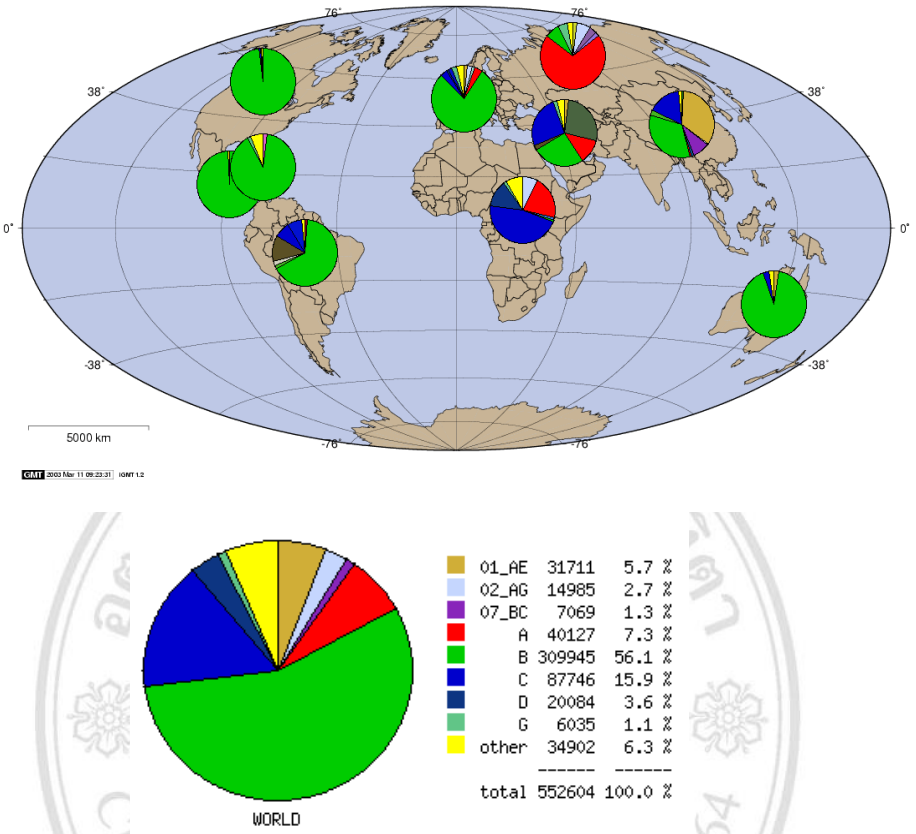

Figure 2.7 The global distribution and prevalence of HIV-1 subtypes and recombinant forms (based on the subtype distribution available from the Los Alamos HIV Database; http://www.hiv.lanl.gov/).

In Asia, national HIV prevalence is highest in South-East Asia. The studies of Hamelaar J, et al. 2006 showed the global genetic diversity of HIV in South-East Asia, where 84% of all HIV infections are CRF01_AE. Subtype B and other recombinants reported about 4%. So this region has the combination of CRF 89%, the highest in the world. In Thailand, Cambodia and Viet Nam, CRF01_AE is spread more than 95% of infections, follow by subtype B and other recombinants. In Myanmar, CRF01_AE is found 52% of infections, along with subtype B (24%), C (12%) and other recombinant (12%) [72].

In China, CRF07_BC (35.5%), CRF01_AE (27.6%), CRF08_BC (20.1%), and Thai subtype B or B' (9.6%) were the four main HIV-1 strains in various risk groups between 2006 and 2008. CRF07_BC and CRF08_BC were predominant infection

among IDU. However, MSM is the risk group of HIV infections within China, and relate with different subtypes of HIV-1 infection and CRFs [136-140]. Western Yunnan, bordering with Myanmar, is a known hotspot of recombination includes CRF07_BC (18.9%), CRF08_BC (39.1%), and CRF01_AE (22.4%) [141]. In Dehong, Yunnan, pol sequences in newly diagnosed HIV-infection showed subtype C accounted for 43.1%, URF for 18.4%, CRF01_AE for 17.7%, B for 10.7%, CRF08_BC for 8.4%, and CRF07_BC for 1.7% [142]. In Kunming, Yunnan, multiple genotypes were identified among MSM, including CRF01_AE (64.9%), CRF07_BC (25.2%), URFs (5.3%), subtype B (3.1%) and CRF08_BC (1.5%). The mean of genetic distance within CRF01_AE were larger than that within CRF07_BC [143].

In order to estimate the subtype viability of HIV-1 among IDUs in Laza and Maizayang of northern Myanmar during June to August 2009, Pang and his team analyzed sequencing of the p17, pol, vif-env and C2V3 fragments in 83 individuals [144]. They found a significantly high amount (86.1%) of HIV-1 intersubtype recombinants, and a very low incidence of subtypes B' (3.8%), C (7.6%) and CRF01_AE (1.3%) in these HIV-infected IDUs. The four recombinant patterns formed were CRF01_AE, B and C. The most dominant recombinants were B/C and CRF01_AE/B/C, accounting for 54.4% and 42.6% of all cases, respectively. When comparing with well-known CRFs or URFs, these recombinants had different patterns from each other; forming 64 URFs. This study pointed to the very broad complexity of HIV-1 recombination among CRF01_AE, B' and C in northern Myanmar, which was never previously reported in Asia [144]. In addition, northern Myanmar is the center of trade, tourism and travel between bordering countries, and had reported HIV-1 subtype B, C and CRF01_AE from this area that had developed from Chinese and Myanmar strains [21].

In the Philippines, Paladin and his colleagues studied the genetic variability in 51 HIV-1-positive Filipinos infected from 1987 to mid-1996. This was done by the sequencing of a 204 base-pair fragment of the *env* C2-V3 region, and analyzed phylogenetic relationships among the DNA sequences. The results showed the five subtypes composing of subtype B (n = 37), subtype E (n = 8), subtype A (n = 3), subtype C (n = 2) and subtype D (n = 1), with a difference of *env* nucleotides

ranged from 11.7 to 32.2%. Subtype B, C and some subtype E did not show a genetic relationship to Asian sequences that caused by *env* mutational escape of ARV. The report of genetic diversity in South and South-east Asia might change because about two-thirds were infected with HIV-1 outside the Philippines. The mode of transmission in almost all subjects was sexual transmission (94%), however it could not classify the subtype by mode of transmission [145].

In Indonesia, the majority of people living with HIV were infected by IDUs and sexual transmission. In 2007, Sahbandar and his staff studied molecular epidemiology profiles in both groups involving 208 individuals in Jakarta. Phylogenetic analysis of *gag* and *env* C2V3 regions showed almost all samples were CRF01_AE (n = 200, 96.2%), and only 3 (1.4%) were subtype B. Five samples (2.4%) referred to recombinant forms that were composed of unique CRF01_AE/B recombination and one sample was CRF33_01B, reported in Malaysia. This study concluded CRF01_AE was predominant in every risk group, that also the pure subtype B decreased, and the first report of CRF01_AE/B recombinant HIV was found in Indonesia [146]. In 2012, Merati and his colleagues reported that 108 people with HIV-1 subtypes in Indonesian were identified as 65 IDUs and 43 sexual route infections. After analyzing the *env* region, it showed 4 subtypes that consisted of CRF01_AE (n=96, 88.9%), B (n=10, 9.3%), C (n=1, 0.9%) and G (n=1, 0.9%). Moreover, 100% of IDUs infected were with CRF01_AE, while subtype B was more common among sexually acquired infection [147].

For the study in Thailand, Subbarao and his team investigated 214 asymptomatic HIV-1 infected people with various modes of transmission from nine provinces that composed of Chiang Mai, Chiang Rai, Phitsanulok, Khon Kaen, Ubon Ratchathani, Nakhon Ratchasima, Trang and Songkhla during 1994 and 1995. HIV-1 subtypes and their genetic diversity were verified by a combination of direct DNA sequencing (n = 95), subtype-specific oligonucleotide probe testing (n = 201), and V3-loop peptide enzyme immunoassay (PEIA) (n = 214). The results showed *env* subtype E (175; 81.8%) and B (39; 18.2%) that divided to B' (37; 94.9%) and North American-like B strains (BZ) 2 (5.1%). Most subtype B strains found in Thailand are part of a distinct sub cluster within the subtype B branch on phylogenetic trees, termed

B'; formerly Thai B or BB. In 149 people with sexual risk behaviors could classify to subtype E (146, 98.0%). 65 viruses found from IDUs, 29 (44.6%) were subtype E and 36 (55.4%) were subtype B, consisting of 35 B' strains. In IDUs, there was a regional variant in the amount of subtypes E and B'. The intrasubtype nucleotide within the V3 and flanking *env* gene had a low viability (5.7% for subtype E and 3.1% for subtype B') compared with other subtypes from different countries that indicated Thailand may be a good choice for the evaluation of HIV-1 vaccines [8].

To identified HIV-1 genotypes in candidate populations for a prime-boost phase III vaccine trial in Rayong and Chon Buri Province, Thailand that collected from 1998 to 2001. A new multiregion hybridization assay, MHAbce was applied to complete genome sequencing and distinguishing HIV-1 subtype. 168 of 194 HIV-1 infections were genotyped. These results indicate that incident and prevalent HIV-1 infections in this populations were 90% CRF01_AE, 3% subtype B, and 7% either recombinant or dual infection. Two CRF01_AE/B and one CRF01_AE/C recombinant were identified among selected prevalent infections [148].

In the RV144 HIV vaccine trial. Thailand, among 390 HIV-1 infection volunteers who were deferred from enrollment showed the subtype distribution: CRF01_AE9 1.7%, subtype B 3.5%, B/CRF01_AE recombinants 4.3%, and dual infections 0.5%. CRF01_AE strains were 31% more diverse than those from the 1990s Thai epidemic that informed vaccine immunogen design [149].

The study of HIV-1 genetic diversity in recently seroconverting (<12 months) Thai repeated blood donors attending the National Blood Centre, Thai Red Cross Society (NBC, TRCS) from September 2007 until March 2008 was assessed. Ten HIV-1 recent seroconvertors were identified six CRF01_AE, one subtype B, and two unique CRF01_AE/B recombinants. The estimated median time to seroconversion was 67.3 days (range: 45.5-102.0 days), and viral load ranged from 307 to 341,805 copies HIV-1 RNA/ml [10].

Sirirungsi and her colleagues wanted to know the results of a virus diversity and genetic recombination of HIV-1 viruses in the next population, after knowing the HIV epidemic in Thailand was initially dominated by CRF01_AE in heterosexual

transmissions and followed by subtype B among IDUs. Investigations of HIV-1 strains among HIV infected pregnant women. The study involved 752 infected pregnant women enrolled in three different period trials of perinatal HIV-1 prevention over 12 years, PHPT-1 (1997-1999, 300 women), PHPT-4 (2004-2005, 239 women) and PHPT-5 (2009 - 213 women). HIV serotypes can be classified by a competitive indirect ELISA named subtype-specific enzyme immunoassays (HIV-1 SSEIA) [150] based on V3 peptides of B and E subtypes. The results showed subtype E in three trials was 77%, 62% and 67%, respectively while subtype B accounted for 11% in the first, 25% in the second and 8% in the third. Moreover, the increase in the proportion of indeterminate serotypes by SSEIA was 12%, 13% and 25% in the three trials, respectably, because of low serum reactivity. This reactivity may reflect the lowering of V3 specific antibodies or a more recent infection. HIV-1 serotype E had a high prevalence in Thailand, however it had decreased over time [151].

To further investigate HIV-1 molecular epidemiology in northern Thailand in 2011, Praparattanapan and her team performed in-house genotypic assays in 420 samples from treatment-experienced patients. The results suggested that HIV-1 CRF01_AE was still predominant (94.8%), followed by subtype B (2.9%) and C 2 samples (0.5%) that patients were migrant and who received HIV-1 subtype C outside Thailand by sexual transmission. Moreover they found the intersubtype recombination of CRF01_AE with subtype B or subtype C (0.71%) [152].

Based on the previous data, monitoring the evolution of HIV-1 diversity in Thailand is needed to gauge the efficacy of HIV prognosis, diagnosis, antiretroviral treatment monitoring, prevention and vaccine design. This study aims to investigate HIV-1 subtype distribution and intersubtype recombinants circulating in Northern Thailand. This region maintains high levels of infection, while also being heavily traveled to for trading and traveling. Along with Thailand's participation in AEC and AFTA in 2016, the moving of people to the region will directly effect to the genetic diversity of HIV in Thailand.

Objective

1. To understand of HIV-1 epidemiology in Northern Thailand
2. To gain beneficial insight into genetic characterization in recombinant HIV-1 infected patients in Northern Thailand