

UNIVERSITY OF HAWAII LIBRARY

BIOINFORMATIC APPROACH FOR
TRACKING HIV-1 EVOLUTION IN VIETNAM AND
NEIGHBORING SOUTHEAST ASIAN COUNTRIES

A THESIS SUBMITTED TO THE GRADUATE DIVISION OF THE
UNIVERSITY OF HAWAI'I IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

IN

BIOMEDICAL SCIENCES
(TROPICAL MEDICINE)

By

Fumiyuki Isami

Thesis Committee:

Vivek R. Nerurkar, Chairperson

Sandra P. Chang

Allison Imrie

Abstract

There is high prevalence of a methionine substitution at the tip of the V3 loop (MGPGQ) among HIV CRF_AE strains from Vietnam. The aim of this study was to identify other molecular markers or “signature sequences” for mapping the spread of HIV in Vietnam and in neighboring Southeast Asian countries.

Analysis of the sequence diversity and grouping by molecular markers suggested that ET-strain initially gained entry in CSW in southern Vietnam. Unique substitutions among ECM and EC- strains in southern Vietnam IDU and CSW suggested independent introduction and spread of HIV among these high-risk groups. Unique and identical amino acid substitutions found in ECV strains from IDU in northern Vietnam and southern China suggested cross-border travel of virus-infected IDU.

TABLE OF CONTENTS

Abstracts.....	iii
List of Tables	vii
List of Figures	viii
Chapter 1. Introduction	1
Bioinformatics.....	1
Subtype of HIV-1.....	2
HIV Genes.....	4
HIV in Southeast Asia	5
Chapter 2. Methods.....	10
Study Population.....	10
RNA Extraction	12
RT-PCR.....	13
Amplification of HIV-1 env gene.....	14
Subtyping.....	15
Hypermutation and syncytium inducing (SI) viruses.....	15
Mean p-distance	18

Minimum Evolution Tree	18
V3 motifs and N-linked glycosylation sites.	19
Viral Epidemiological Signature Pattern Analysis (VESPA).....	20
Synonymous and Non-synonymous Distance (dN/dS).....	20
Evolutionary Rate.....	21
Chapter 3. Results.....	23
Study Population.....	23
Subtyping.....	23
Hypermutation and SI viruses	24
Mean p-distance	27
Minimum Evolution Tree	29
V3 motifs and N-linked glycosylation sites	32
Viral Epidemiological Signature Pattern Analysis (VESPA).....	33
Synonymous and Non-synonymous Distance (dN/dS).....	38
Evolutionary Rate.....	40
Chapter 4. Discussion.....	41

Epidemiological Information.....	43
Phenotypical Information	46
Subtype	46
Cell Tropism	47
Viral differences between transmission pathway (IDU and CSW).....	49
References.....	52

LIST OF TABLES

<u>Table</u>		<u>Page</u>
1.	Study Population	11
2.	Hypermuted and syncytium inducing (SI)	27
3.	The tetrapeptide motif of the V3 loop.....	34
4.	Viral Epidemiological Signature Pattern.....	35
5.	Definition of four variant subtypes by VESPA method	36
6.	The dN/dS ratio.....	39
7.	The absolute rate of molecular evolution.....	40

LIST OF FIGURES

<u>Figure</u>		<u>Page</u>
1.	Nucleotide and Amino Acid Sequence of HXB2 (Subtype B) and CM240	17
2.	Overall mean p-distance by category.....	29
3.	A Minimum-Evolution tree was constructed by 402 CEF02_AE sequences.....	31
4.	Distribution of the variant subtypes grouped by minimum evolution and VESPA methods in areas and risk groups.....	37

Chapter 1. Introduction

Bioinformatics: DNA sequencing has made it us possible to analyze genes at the nucleotide level (73). First few researchers could access this technology, as the process was complicated, slow, and costly. As methods were improved with the invention of polymerase chain reaction (PCR) (44) and other modern techniques, sequencing methods became more widely available. The increase of sequence data generated another problem: how to manage the numerous data sets. To solve this problem, worldwide sequence databanks, such as GenBank, were established.

In addition to basic sequence data, gene structure and function needed to be organized, and the relevant databanks were established. However, individual researchers still need to know how to manage the databanks, for example, how to retrieve and compare necessary information from the exponentially increasing data sets.

Under such circumstances, the demand for bioinformatics is increasing. The term, Bioinformatics, is defined as “the science of using information to understand biology.” Strictly speaking, bioinformatics is a subset of the larger field of computational biology, the application of quantitative analytical techniques in modeling biological systems (18). Phylogenetic analysis is

one such application in bioinformatics, and molecular epidemiology is a tool to track the evolution of microbes and their host.

Subtypes of HIV-1: The simian immunodeficiency virus (SIV) circulates among primates. SIV entered the human population through multiple zoonotic infections from non-human primates (6). Based on molecular phylogenetic data, transmission from non-human primates to humans occurred in the 1930's (26, 72). In fact, the virus was isolated from a plasma sample taken in 1959 from an adult male living in what is now the Democratic Republic of Congo (6). This strain was pathogenic to human although at first, SIV was non-pathogenic in its human host.

In 1981, The United States Centers for Disease Control and Prevention (CDC) observed an unusually high number of requests for the drugs used in the treatment of *Pneumocystis carinii* pneumonia (PCP) and Kaposi's Sarcoma (KS). A new clinical syndrome was initially called GRID (gay-related immune deficiency), was soon recognized in homosexual males. The disease was characterized by immunodeficiency resulting from severely reduced CD4+ T lymphocyte counts. Epidemiological and in-vitro investigation led to the hypothesis that the cellular immunity dysfunction that predisposed the affected individuals to opportunistic infections and tumor was resulted to a common exposure. The term, AIDS (acquired immunodeficiency

syndrome) was coined and defined in 1982. AIDS cases were soon reported in other populations, including injection drug users, heterosexual men and women, and hemophiliacs.

In 1984, two independent groups from the United States and France isolated a retrovirus from the lymph node of a patient at risk of developing AIDS. This virus was eventually classified as a lentivirus and named human immunodeficiency virus, (HIV)

The HIV pandemic has severely impacted populations around the globe. Within 20 years, HIV had infected 58 million people and killed 22 million of them (85). Even though anti-viral therapy has dramatically decreased the mortality rate among infected patients, the HIV pandemic is still growing.

Due to the high error rate of reverse transcriptase, HIV is prone to substitutions (41, 88), recombination (65), insertions, and deletions. Furthermore, the average HIV production rate is estimated to be 10.3×10^9 virions per day (57). The HIV-1 *env* gene region, which is important for virus entry in to the cell, is highly mutated; the evolutionary rate is 0.0024 substitutions per base pair per year (26). The very high rate of HIV mutation has made vaccine design difficult.

These unique HIV properties allow us to categorize HIV in to different subtypes (28, 64). There are three main groups: M (for main), N (for non-M, non-O), and O (for outlier). The group M dominates the HIV pandemic, of which subtypes are A, B, C, D, F, G, H, J, and K. Newly

discovered subtypes are named alphabetically. Subtypes E and I are problematic and often found as a recombinant form. Subtype E is considered as a recombinant between A and E, although no full-length subtype E representatives have been found. Hence, subtypes E and I were renamed CRF01_AE and CRF04_cpx respectively; CRF stands for *circulating recombinant forms*. However, another study has suggested that there is inadequate support for the hypothesis that CRF01_AE variants were derived from a recombinant lineage.(4).

Subtype B is predominant in North America, South America, Europe, Australia, and Far East Asia. Subtype C dominates the South African and Central Asia HIV pandemic, and CRF01_AE causes severe outbreak in Southeast Asia. Subtype F is emerging in some parts of Africa including Ethiopia and India (1) (<http://hiv-web.lanl.gov/cgi-bin/geography>).

HIV subtypes can be characterized based on heteroduplex mobility assay and by sequencing. Molecular epidemiological studies have often employed subtype analysis to trace routes of transmission and migration. Additionally, by using sequence data, viruses can be sub-classified within the same subtype. Molecular epidemiology is routinely used to estimate the migration pattern and the age of the pandemic.

HIV genes: The three major structural genes encoded by HIV are: *gag*, *pol*, and *env*. The *gag* gene product is a capsid protein, and *pol* encodes the polymerase gene. The *env* gene encodes the viral

transmembrane region and is frequently mutated; this has allowed the virus to evade the hosts' acquired immune response. Traditionally, the *env* protein has been used for epidemiological study of retroviruses. Even prior to the discovery of HIV, the envelope antigen of a retrovirus called, the human T-cell leukemia virus was used for epidemiological studies (20, 33, 53). Consequently, the HIV *env* gene, specifically spanning the V3 loop region, has been extensively used to study the migration of HIV.

We sequenced the *env* gene spanning the V3 gene region as well, because this region was intensively studied for epidemiological purposes, and we could compare our data with the sequence data submitted by other groups (10, 30, 40, 42, 47-49, 54, 59, 74, 79, 83, 86, 87, 90, 91). Another reason was that the *env* gene region was prone to the mutations, and slight differences could be recognized. We can interpret the results more clearly based on the *env* gene when compared to other genes.

HIV in Southeast Asia: The history of HIV in Southeast Asia is relatively young. The first HIV-1 infection and AIDS was reported among Thai male prostitutes and thalassemia patients in the mid-1980s (58, 84). However, a sudden rise of HIV-1 infection was noted during the year 1998 among injection drug users (IDU) in Bangkok (86). In 1989, the first national sentinel

seroprevalence survey detected a 3.5% median among brothel-based female prostitutes in the Thai province, with the highest rate of 44.0% in Cheng Mai Province (82). In 1991, the World Health Organization (WHO) selected Thailand as a potential HIV vaccine evaluation site. By the end of 1999, 755,000 people are living with HIV/AIDS, and 128,606 AIDS cases were reported (81).

The epidemiology of HIV-1 in Thailand was unique to other countries. In the early phase of the pandemic, subtype B was predominant among IDU. Subtype B is the dominant strain in the western hemisphere, and HIV in Thailand was genetically similar to the subtype B viruses found in the Americas and in the Europe prior to 1988. However, the late subtype B strains identified in Thailand became genetically different from typical subtype B. Therefore, the Thai subtype B strains have been referred to as B', Thai, or BB (21, 23, 40, 54, 55).

In the later phase, CRF01_AE appeared among commercial sex workers (CSW) and heterosexuals and ultimately took over subtype B' strains (23, 40, 55). CRF01_AE originated from Africa, and the estimated year that a single strain was introduced into Southeast Asia was around 1986 (26, 41, 45). This strain appeared in the northern part of Thailand even though subtype B appeared initially in Bangkok and Southern Thailand. Surveillance during 1994 and 1995 revealed that the proportion of CRF01_AE in Thailand reached 81.8% (78).

The emergence of HIV-1 in Thailand has severely impacted other Southeast Asian countries, such as Cambodia, Myanmar, Malaysia, Vietnam, Indonesia, Singapore and Southern China. In addition, countries such as Japan (89), Korea (32), Singapore (22), Taiwan (75) and European countries (34, 38, 56, 60, 61, 63, 76, 77), which have many travelers and visitors to Southeast Asia, were also impacted.

In Cambodia, the first HIV-1 case was detected in 1991. There is little evidence for transmission by IDUs but spread has been through heterosexual contacts. Prostitution played the most influential factor for this epidemic. Seroprevalence rates among CSWs were as high as 39.3% according to the 1998 UNAIDS report. All genotyped viruses clustered with HIV subtype CRF01_AE (30).

Myanmar (Formerly known as Burma) is surrounded by three countries: India where subtype C is circulating, and Thailand and Yunnan Province of China where CRF01_AE and B' are circulating. As a result, predominant strains varied with the geographic area. Subtype B' is predominant in Yangon, the capital city of Myanmar; however, CRF01_AE was observed in a small IDU population. The cities bordering Thailand had a CRF01_AE HIV epidemic. In central and northeast Myanmar, viruses were distributed with mixture of CRF01_AE and subtype B' (31).

Other surveillance reported the existence of subtype C and other recombinants such as C/B' and C/B'/E in Mandalay, central Myanmar (43).

HIV-1 was introduced in to Vietnam slowly in the early 1990's. Only a few cases were reported by the end of December 1991. Ten cases were detected in the following year, which led to mistaken speculation that AIDS would not be a major health problem in Vietnam. In 1993, sentinel surveillance detected an HIV outbreak among IDU in the south; the total number of cases that year reached 1,148. Since 1993, there has been a rapid increase in HIV infections (48). In 2002, medical checks of young army recruits in a southern Vietnamese province found that 4.5 percent were HIV-positive (2). This report suggested that HIV was not only limited to high-risk populations such as IDU and CSW, but also was present in the general population. The rate of infection among CSW was 1.5% in 1999 but increased to 3.53% in 2000. As of November 30, 2002 there were 42,365 HIV cases, 6,343 of which were classified as AIDS, and 3,474 died (2).

Although HIV prevalence rates varied greatly by province and risk group in Vietnam, in most provinces, the occurrence of HIV was highest among IDU: 65% of Vietnam's HIV infection (14). However, in An Giang, the province bordering Cambodia, the occurrence of HIV-1 was the highest among CSW relative to the rest of the country (25). The epidemic pattern of HIV in

Vietnam may be similar to that which occurred in Thailand based on the similarity in nucleotide sequences. Their strains are mainly CRF01_AE.

However, Vietnam strains have some distinguishable characteristics, which may be caused by a founder effect. The previous study utilized the viral epidemiological signature patterns for mapping the HIV-1 spread in Vietnam. Methionine substitution just before the crown tetrapeptide GPGQ motif was observed frequently in HIV-1 CRF01_AE in both IDU and CSW in South Vietnam (25, 48). This substitution could be a signature pattern to identify HIV-1 strains in Vietnam. It was reported that the valine substitutions, which occurs 12 amino acids downstream of the V3 loop, were observed only among the IDU and CSW in the North Vietnam and in South China, suggesting circular migration between two countries (24, 25).

Therefore, the aim of this study was to map the spread of HIV-1 in Vietnam, by utilizing bioinformatics methods, such as nucleotide distances, viral epidemiological signature pattern and the phylogenetic tree method.

Chapter 2. Methods

Study Population: Specimens were collected from high-risk populations such as CSW, IDU, and their partners, patients attending sexually transmitted disease (STD) clinics, and prisoners. For this study, conducted between, 1995 to 1999, the surveillance spanned four provinces in Northern Vietnam and six provinces in Southern Vietnam. Additionally, we collected plasma samples from AIDS patients whose risk factors for acquisition of HIV were unknown. Epidemiological data such as risk factor(s) for acquisition of HIV, age, gender, and marital status, was collected from the patients. It was reviewed by the institutional human subject review board (IRB) both at University of Hawaii and in Vietnam, conducted in accordance with federal regulations. The recruitment of volunteers has been done with appropriate procedures for obtaining informed consent of study subjects.

A total of 240 plasma samples were collected from HIV-1 infected patients: 51 from North Vietnam (45 IDU, 3 CSW, 1 STD, 1 Prisoner, and 1 IDU partner), and 189 from South Vietnam (77 IDU, 72 CSW, 22 STD, 2 IDU partner, 5 self identified both IDU and STD, and 11 unknown).

TABLE 1: Study Population

Area*	IDU**					CSW**					
	1995	1997	1998	1999	Total	1995	1996	1997	1998	1999	Total
North Vietnam (n)											
Langson (6)	2		4		6						
Hanoi (17)	5		8	6	19						
Quang Ninh (13)			8	5	13						
Hai Phong (13)			6	4	10				3		3
Total (51)	7		26	15	48				3		3
South Vietnam (n)											
An Giang (76)	10	6	3	4	23	13	2	13	9	16	53
Da Nang (15)		7	5	2	14				1		1
Can Tho (17)			6		6	2			7	2	11
Kien Giang (9)			7		7				2		2
Ho Chi Minh (45)	4	4	15	10	33	3		1	4	4	12
Nha Trang & Khan Hoa (27)	3	2	8	9	22				3	2	5
Total (189)	17	19	44	25	105	18	2	14	26	24	84

* The HIV seropositive patients were recruited from 4 provinces in North Vietnam and 6 provinces in South Vietnam. The parenthesis by the province is the number of HIV positive patients.

** The patients were categorized by their risk factor of Injection Drug User (IDU), Commercial Sex Workers (CSW). Prisoner, STD patients, and unknown were re-categorized to IDU or CSW according to their gender. IDU partner was re-categorized to IDU. The sample collection year was shown.

Table 1 showed the number of the collected samples in 10 provinces: the provinces of North Vietnam included Langson (6 IDU), Hanoi (17 IDU, 1 STD, and 1 prisoner), Quang Ninh (13 IDU), and Hai Phong (9 IDU, 3 CSW, and 1 IDU partner). The provinces of South Vietnam included Da Nang (14 IDU and 1 self identified STD and IDU), An Giang (16 IDU, 45 CSW, 5 STD, and 10 unknown), Can Tho (3 IDU, 10 CSW, and 13 STD), Kien Giang (6 IDU, 2 CSW, and 1 unknown), Ho Chi Minh (21 IDU, 10 CSW, 13 STD, 1 and IDU partner), and Khanh Hoa (17 IDU, 5 CSW, 4 self identified STD and IDU, and 1 IDU partner). The prisoners, STD patients, and unknown were re-categorized in to IDU or CSW based on their gender. IDU partners were

re-categorized to IDU. After re-categorization, there were 48 IDU and three CSW in North Vietnam and 105 IDU and 84 CSW in South Vietnam.

RNA extraction: Viral RNA was extracted from the plasma samples. The RNA extraction was done by Qiamap Viral RNA kit (Qiagen inc.) from 140 μ L of plasma. We followed the spin column procedure for the RNA extraction, according to the manufacturer's protocol: (1) Pipet 560 μ L of prepared Buffer AVL containing Carrier RNA into a 1.5-mL microcentrifuge tube, (2) Add 140 μ L plasma to the Buffer AVL/carrier RNA in the microcentrifuge tube. Mix by pulse-vortexing for 15 sec, (3) Incubate at room temperature (15–25°C) for 10 min to lyse viral particles, (4) Briefly centrifuge the 1.5-mL microcentrifuge tube to remove drops from inside of the lid, (5) Add 560 μ l of 100% ethanol to the sample, and mix by pulse-vortexing for 15 seconds, (6) Apply 630 μ l of the solution from the previous step to the QIAamp spin column (in a 2-mL collection tube) without wetting the rim. Close the cap, and centrifuge at 8,000 rpm for 1 min. Place the QIAamp spin column into a clean 2-mL collection tube, and discard the tube containing the filtrate, (7) Open the QIAamp spin column, and repeat the previous step, (8) Carefully open the QIAamp spin column, and add 500 μ L of Buffer AW1. Close the cap, and centrifuge at 8,000 rpm for 1 min. Place the QIAamp spin column in a clean 2-ml collection tube and discard the tube

containing the filtrate, (9) Open the QIAamp spin column, and add 500 μL of Buffer AW2. Close the cap and centrifuge at full speed 14,000 rpm for 3 min. (10) Place the QIAamp spin column in a clean 1.5-mL microcentrifuge tube. Discard the old collection tube containing the filtrate. Carefully open the QIAamp spin column and add 60 μL of Buffer AVE equilibrated to room temperature. Close the cap, and incubate at room temperature for 1 min. Centrifuge at 8,000 rpm for 1 min. The viral RNA was stored at -80°C till further use.

RT-PCR: Eight μL of viral RNA was employed to synthesize cDNA using random hexamers and Superscript II H- enzyme (Life Technologies, Grand Island, NY). Based on the manufacturer's instructions, 0.5 μL Rnase Inhibitor, 0.5 μL Random Hexamers, and 8.0 μL viral RNA was dispensed into a reaction tube. The viral RNA was denatured on a heating block at 70°C for 10 min, and kept on ice 2 min. The other reagents were added into the tube, such as 4.0 μL RT Buffer, 2.0 μL DTT, 4.0 μL dNTP, and 1.0 μL SuperScript. The cDNA synthesis was conducted by placing the tube at 42°C for 50 min, followed by 70°C for 15 min. The cDNA was stored at -30°C till further use.

Amplification of HIV-1 env gene: PCR was conducted to amplify a 330-bp *env* gene region spanning the V3 loop (7007 to 7337 in HXB2 numbering). Oligonucleotide primers were derived from the sequence alignment of HXB2 strain. One pair of primer for outer PCR was used to amplify the *env* gene: KK40 (5'-ACAGTACAATGTACACATGG-3') and KK 80 (5'-CCACTCTTCTCTTTGCCTTGGTGGG-3'). The primers for the nested PCR were R124 (5'-AAATGGCAGTCTAGCAGAAG-3') and KK30 (5'-AATTTCTTGGTCCCCTCCTG-3'). All PCR negative specimens using the above primer pair were re-amplified by using env10 primer (5'-GATGGGAGGNGCATAACAT-3') instead of EKK30, and the amplicon size was 534 bp (7007 to 7541 HXB2 numbering).

These primers were used at a concentration of 0.1 mM in a reaction mixture of 25 μ L containing 2.5 μ L of 10 mM Tris-HCl (pH 8.3), 1.8 μ L of MgCl₂, 2 μ L of each dNTP, 0.125 μ L of each primer, 0.125 μ L of *Thermus aquaticus* DNA polymerase (Perkin-Elmer), 13.3 μ L of DEPC water, and 5 μ L of template. Using a 9600 DNA thermal cycler (Perkin-Elmer), the mixture was initially denatured at 95°C for 1 min., then cycled 10 times at 95°C for 15 sec., 45°C for 30 sec., 72°C for 90 sec., followed by 35 cycles at an annealing temperature at 50°C, with a final extension of 7 min. at 72°C before storing at 4°C.

For nested PCR, the reaction mixture contained 6.75 μ L of 10 mM Tris-HCl (pH 8.3), 3.6 μ L of MgCl₂, 4 μ L of each dNTP, 0.5 μ L of each primer, 0.25 μ L of *Thermus aquaticus* DNA polymerase (Perkin-Elmer), 29.4 μ L of DEPC water, and 5 μ L of template. The thermal cycle conditions for PCR amplification were same as for the outer PCR. The amplicons were sequenced using an automated sequencer (model 373A, Applied Biosystems).

Subtyping: We employed HIV-1 genotyping tool by BLAST algorithm (<http://www.ncbi.nlm.nih.gov/retroviruses/subtype/subtype.html>)(3). To confirm the genotyping results, we constructed a minimum-evolution tree (71) based on our sequences and other reference sequences defined by the Los Alamos HIV sequence database.

Hypermutation and syncytium inducing (SI) viruses: Using the program HYPERMUT (69), we identified the hypermutated sequences, using CM240 (U54771) as an ancestor HIV sequence, which was amplified from an asymptomatic heterosexual 21-year-old Thai man in 1990. We defined hypermutated sequences as sequences with more than 10% Gs: the percentage of Guanines in the reference sequence that has undergone Guanine -->Adenine transitions.

Another method was to estimate the non-syncytium inducing (NSI) and syncytium inducing (SI) viruses. The SI and NSI viruses were defined genetically based on the presence or absence of arginine or lysine at amino acid positions 301, 303, 306, 308, 315, or 322 (5, 39). FIG 1 shows the position number of both amino acid and nucleotide base, based on HIV Sequence Locator Tool of Los Alamos HIV sequence database (7).

Both hypermutation and SI-NSI classification could indicate the heavily mutated sequences. The objective of this analysis was to view the distribution of the viruses that had selective pressure such as antiviral therapy or long host's immunity. Consequently, we can find the relative time of HIV introduction in the study groups.

The results of this analysis were used to minimize the error of the estimation for the evolutionary rate and Viral Epidemiological Signature Pattern Analysis discussed later.

	7050	7060	7070	7080	7090											
	277	278	279	280	281	282	283	284	285	286	287	288	289	290	291	292
HXB2	---TTTCACGGACAATGCTAAAACCATAAATAGTACAGCTGAAACACATCTGT															
CM240	---CTCACAAACAATGCCAAAACCATAAATAGTGACACCTTAATAAATCTGT															
	I	T	N	N	A	K	T	T	T	V	H	I	N	K	S	V

	7100	7110	7120	7130	7140												
	293	294	295	296	297	298	299	300	301	302	303	304	305	306	307	308	309
HXB2	AGAAATTAATTGTACAAGACCCAAACAACAATACAAGAAAAAGAATCCGTA																
CM240	AGAAATCAATTGTACCAGACCTCCAACAATACAAGAAAGTATAACTA																
	F	I	N	C	T	R	P	S	N	N	T	R	T	S	I	T	I

	7150	7160	7170	7180	7190											
	310	311	312	313	314	315	316	317	318	319	320	321	322	323	324	325
HXB2	TCCAGAGAGGACCAGGGAGAGCATTGTACAATAGGAAAA---ATAGGAAAT															
CM240	TA-----GGACCAGGACGAGTATTCTATAGAACAGGAGATATAATAGGAAAT															
		G	P	G	R	V	F	Y	R	T	G	D	I	I	G	N

	7200	7210	7220	7230	7240												
	326	327	328	329	330	331	332	333	334	335	336	337	338	339	340	341	342
HXB2	ATGAGACAAGCACATTGTAAACATTAGTAGAGCAAAATGGAAATAACACTTT																
CM240	ATAAGAAAAAGCATATTGTGAGATTAATGGAAACAATGGAAATAAGTTTT																
	I	R	K	A	Y	C	F	I	N	G	T	K	W	N	K	V	I

	7250	7260	7269	7280	7290													
	343	344	345	346	347	348	349	350	351	352	353	354	355	356	357	358	359	
HXB2	AAAACAGATAGCTAGCAAATTAAGAGAACAATTTGGAAATAATAAAACAA																	
CM240	AAAACAGGTAACCTGAAAAATTAAGAGACTTT-----AATAAGACAA																	
	K	Q	V	T	F	K	I	K	F	H	F	-	-	-	N	K	T	I

	7300	7310	7320				
	360	361	362	363	364	365	366
HXB2	TAATCTTTAAGCAATCCTCAGGA						
CM240	TAATCTTTCAACCACTCAGGA						
	I	F	Q	P	P	S	G

	7350	7360	7370					
	376	377	378	379	380	381	382	383
HXB2	TTTAATTGTGGAGGGGAATTTTT							
CM240	TTTAATTGTAGGGGGGAATTTTT							
	F	N	C	R	G	F	F	F

FIG. 1. Nucleotide and Amino Acid Sequence of HXB2 (Subtype B) and CM240 (CRF01_AE): High charge at the position in BOLD (301, 303, 306, 308, 315, and 322) indicates SI. The amino acid (309, 340 and 343) and the nucleotide (7269) in the square are epidemiological signature pattern substitutions. The position number was based on the HIV Sequence Locator Tool developed by Los Alamos HIV Sequence Database.

Mean p-distance: There were a total of 402 samples: IDU in North Vietnam (n=48), CSW in North Vietnam (n=3), IDU in South Vietnam (n=103), CSW in South Vietnam (n=82), Thailand (n=135), China (n=17), Cambodia (n=12), Malaysia n= (1), and Myanmar (n=1). If non-CRF01_AE strains were found, they were removed from the alignment. Nucleotide sequences of the *env* gene region were aligned by ClustalW program (80), and the alignment error was visually modified. The aligned sequences consisted of a 266-bp (7053 to 7319 by HIV Sequence Locator in FIG 1.) fragment. However because of deletions and short sequences, after gap stripping, the final product was 200 bp.

The sequence was exported to MEGA2 program (29), and the mean p-distance was calculated with respect to risk factors and geographic area. The model was Kimura 2-parameter method, and both transitions and transversions were computed. Standard Error (SE) was computed by the bootstrap method, of which replicates were 500.

Minimum-Evolution Tree: We removed four subtype B sequences and constructed a minimum-evolution tree (71) based on the aligned 402 CEF02_AE sequences including: IDU in North Vietnam (n=48), CSW in North Vietnam (n=3), IDU in South Vietnam (n=103), CSW in South Vietnam (n=82), Thailand (n=135), China (n=17), Cambodia (n=12), Malaysia n= (1), and

Myanmar (n=1). We used the MEGA2 program for constructing the tree, and the model was the minimum-evolution with Kimura-2-parameter, and both transition and transversion were calculated.

V3 motifs and N-linked glycosylation sites: The amino acid positions of V3 crown tetrapeptide motif were 312 to 315 in HIV sequence locator tool. The predominant pattern of the motif was GPGQ in CRF01_AE, but the other patterns were also seen such as GPGR, GPGH, and GPGK. We counted how many sequences had those patterns in area or risk groups.

Glycosylation sites typically have the motif of N-X-S/T, where N, S, or T represents asparagines, serine, or threonine respectively. X can be any amino acid. CM244 had five glycosylation sites within our sequenced range (position 280, 289, 295, 301, and 334), but an additional glycosylation site started from the position 356 in other CRF01_AE sequences. We studied the prevalence of glycosylation motifs in Vietnam population.

Viral Epidemiological Signature Pattern Analysis (VESPA): Viral Epidemiological Signature Patterns are particular sites in amino acid or nucleic acid alignments of variable sequences that are distinctly representative of a query set of sequences relative to a background set. VESPA (Viral Epidemiological Signature Pattern Analysis) program (27) was employed to identify the signature sequence pattern of each geographic area and risk factor. .

We analyzed the signature sequence pattern for four groups: IDU, CSW from North Vietnam and IDU, CSW from South Vietnam. Twenty sequences were randomly selected from each group after removing subtype B, hypermutated, and SI viruses. Since North Vietnam CSW had only three sequences, these sequences were used. In order to identify the signature patterns of IDU in North Vietnam, we set 20 selected sequences from North IDU as query sequences, and the background sequences were 20 randomly selected IDU and CSW sequences in South Vietnam, and 3 of CSW in North Vietnam. We applied the same procedures to identify the viral epidemiological signature patterns for the other three groups.

Synonymous and Non-synonymous distance (dN/dS): HIV is transmitted by different routes among CSW and IDU. Mucosal barrier gives selective pressure to the virus, and limited quasispecies can enter the host via sexual transmission whereas among IDU, all quasispecies can

enter the host as the virus is inoculated directly into the bloodstream and does not cross a mucosal barrier. Therefore, we hypothesized that dN/dS ratio of IDU is different from CSW. dN is the number of non-synonymous substitutions per non-synonymous site, and dS the number of synonymous substitutions per synonymous site. If dN is over dS ($dN/dS > 1$), the population is affected by the selective pressure.

We employed MEGA2 program to calculate synonymous and non-synonymous substitutions within the V3 loop spanning from 7110 to 7218 (108-bp or 36 amino acids), of which model was modified Nei-Gojobori (jukes-Cantor method) (46). A sequence, which did not have a glycosylation site from 301 to 303, was removed from a group of North Vietnam. In South Vietnam, after removing subtype B sequences, we compared the dN/dS value between the sequences with and without glycosylation site. Then removing subtype B, hypermutated, and SI sequences, we analyzed the dN/dS value for IDU and CSW in group I and IV.

Evolutionary Rate: We estimated the evolutionary rates for group I to IV. We removed all subtype B, SI and hypermutated sequences; hence, 44 group I, 38 group II, 13 group III, and 73 group IV were subject to the estimation. These groups contained both IDU and CSW; therefore,

we selected only CSW in group I (n=28) and only IDU in group II (n=13) and IV (n=56) for further analysis.

A maximum likelihood tree was constructed by fastDNAmI program, of which substitution model was F84 (52). As outgroup, CM240 (U54771; CRF01_AE, from Thailand in 1990) was added. We employed TipDate program to estimate the evolutionary rate, based on the maximum likelihood tree and collected year (62).

Chapter 3. Results

Study Population: All CSW were women and 96.7% (148 of 153) IDU were men. Five women were categorized as IDU, and three of them were partners of IDU male patients. The IDU women were, two from An Giang and from Ho Chi Minh City in southern Vietnam. Therefore, the probability of men being in the IDU group, and women in the CSW group was very high.

IDU group was generally older than the CSW group in South Vietnam. The median age of IDU in South Vietnam was 40 years (range 15 – 56 years) while the median age of CSW was 24 yrs (range 15 – 64). Two women IDU from South Vietnam were 30 and 38 years old, older than median age of CSW. Therefore, age may be the indicator of IDU and STD in South Vietnam. However, IDU in North Vietnam were young, the median age was 26 years old (range 17 – 38 years).

The risk factor, STD, was re-categorized since they may have acquired their infection via parenteral or sexual route. STD patients composed of 16 men and six women, and their median age was 34 (range 19 – 47 years) for men and 23 years (range 20 – 27) for women.

Subtyping: Sequences and phylogenetic analysis confirmed that CRF01_AE was the predominant genetic subtype in Vietnam: 236 CRF01_AE (98.3%) vs. 4 subtype B (1.7%). Three of four

subtype B patients were from An Giang: 2 CSW from An Giang and one male AIDS patient with unknown risk factor, who was later re-categorized to IDU. The other was 23-year old male STD patient from Ho Chi Minh City who was also re-categorized to IDU.

We facilitated BLAST® (Basic Local Alignment Search Tool) program to compare 4 subtype B sequences (3). All of them were genetically related to the U.S. subtype B rather than Thai B', suggesting that these patients were infected as a result of foreign contacts.

Hypermutation and SI viruses: Table 2A shows the number of hypermutation and SI viruses in risk- or geographic- groups. We have identified 46 hypermutations (11%) and 154 SI (38%) in 406 sequences. Thirty-one of them had both hypermutation and SI.

We observed hypermutations in the following populations: 2 (4%) in IDU from North Vietnam, 3 (2%) in IDU and 11 (13%) in CSW from South Vietnam, 26 (19%) in Thailand, and 4 (100%) in subtype B. There were 154 SI viruses, which could be classified into 3 (6%) IDU from North Vietnam, 29(28%) IDU from South Vietnam, 26 (31%) CSW from South Vietnam, 89 (65%) from Thailand, 3(25%) from Cambodia, and 4 (100%) in subtype B. The sequences, which had both hypermutation and SI, were from North Vietnam IDU 1 (2%), South Vietnam CSW 5 (6%), Thailand 21 (15%), and subtype B 4 (100%).

In South Vietnam, CSW had more hypermutations than IDU did ($p < 0.05$). However, there was no statistical significance in SI distribution between IDU and CSW from South Vietnam ($p = 0.6$).

Table 2B shows the relationships in among Vietnam hypermutation, SI, and groups divided by the minimum evolution method, which we will discuss later. Subtype B sequences were removed in this table. The number of hypermutation in group I to IV was 12 (13%), 1 (2%), 1 (6%), or 2 (2%) respectively. We observed 42 (45%) SI in group I, 1 (2%) SI in group II, 1 (6%) SI in group III, and 14 (15%) SI in group IV. Only 6 sequences had both hypermutation and SI genotype, and all belonged to Group I.

Group I had hypermutations more frequently than groups II ($p = 0.06$), group III ($p = 0.48$), or group IV ($p < 0.05$). Group I also carried more SI viral strains than group II, group III, or group IV with statistically significance ($p < 0.05$).

Furthermore, we investigated the genetic characteristics of AIDS patients. Theoretically, AIDS patients should carry hypermutation or SI viral strains more frequently than asymptomatic patients. Two of 10 AIDS patients (20%) had hypermutated viruses, and half AIDS patients had SI (5/10). These numbers were higher than average, which were 20% in AIDS vs. 6% in average among hypermutation, and 50% in AIDS vs. 24% in average among SI genotypes in

Vietnam. However, they were not statistically significant ($p=0.12$ for hypermutation and $p=0.07$ for SI viral strains).

On the other hand, there was a significant relationship between hypermutated and SI sequences. Twenty percent of SI viral strains (31 of 154) were hypermutated; on the contrary, 6% of NSI (15 of 252) were hypermutated ($p<0.05$).

TABLE 2A: Hypermutated and syncytium inducing (SI)

	(n)	Hypermutation*	SI**	Both***
<u>North Vietnam</u>				
IDU	48	2(4%)	3(6%)	1(2%)
CSW	3			
<u>South Vietnam</u>				
IDU	103	3(2%)	29(28%)	
CSW	82	11(13%)	26(31%)	5(6%)
China	17			
Thailand	135	26(19%)	89(65%)	21(15%)
Cambodia	12		3(25%)	
Myanmar	1			
Malaysia	1			
Subtype B	4	4(100%)	4(100%)	4(100%)
<u>Total</u>	<u>406</u>	<u>46(11%)</u>	<u>154(38%)</u>	<u>31(8%)</u>

TABLE 2B: Hypermutated and syncytium inducing (SI)

Group****	(n)	Hypermutation	SI	Both
Group I	92	12(13%)	42(45%)	6(6%)
Group II	40	1(2%)	1(2%)	
Group III	15	1(6%)	1(6%)	
Group IV	89	2(2%)	14(15%)	
<u>Total</u>	<u>236</u>	<u>16(6%)</u>	<u>58(24%)</u>	<u>6(2%)</u>

*Hypermutation had over 10% Gs by HYPERMUT program.

**Syncytium inducing viruses were determined by the presence of arginine or lysine at amino acid positions 301, 303, 306, 308, 315, or 322.

***Both meant having both hypermutation and SI.

****Group I to IV were only Vietnam CRF01_AE sequences classified by the minimum evolution method

They were sorted by area and risk factor (TABLE 2A) or groups by minimum evolution (TABLE 2B).

Mean p-distance: The genetic diversity was analyzed using mean p-distance with respect to geographic area and risk factors. The mean p-distance of IDU in North Vietnam was low: 0.039

(n=7, standard error or SE, 0.009) in 1995, 0.035 (n=26, SE, 0.006) in 1998, and 0.024 (n=15, SE, 0.005) in 1999.

CSW in South Vietnam had higher diversity than other risk groups in Vietnam: 0.058 (n=18, SE, 0.008) in 1995, 0.083 (n=2, SE, 0.021) in 1996, 0.078 (n=14, SE, 0.012) in 1997, 0.077 (n=26, SE, 0.009) in 1998, and 0.087 (n=24, SE, 0.01) in 1999.

The genetic diversity of IDU in South Vietnam was moderate: 0.052 (n=17, SE, 0.007) in 1995, 0.060 (n=19, SE, 0.007) in 1997, 0.061 (n=44, SE, 0.007) in 1998, and 0.063 (n=25, SE, 0.008) in 1999.

To compare the diversity with other neighboring countries, we retrieved from the Los Alamos HIV database, HIV sequence data from Thailand, Cambodia, and China.

The p-distance in Thailand in 1990, 1992, 1993, 1994, 1995, and 1996 was 0.047 (n=13), 0.029 (n=7), 0.101 (n=74), 0.072 (n=6), 0.122 (n=18), and 0.102 (n=8), respectively.

FIG. 2 shows the mean p-distance as follows: 0.035 in IDU from North Vietnam (n=48, SE, 0.005), 0.004 in CSW from North Vietnam (n=3, SE, 0.003), 0.067 in IDU from South Vietnam (n=103, SE, 0.007), 0.086 in CSW from South Vietnam (n=82, SE, 0.009), 0.089 in Thailand (n=135, SE, 0.009), 0.058 in Cambodia (n=12, SE, 0.008), and 0.017 in China (n=17, SE, 0.004)

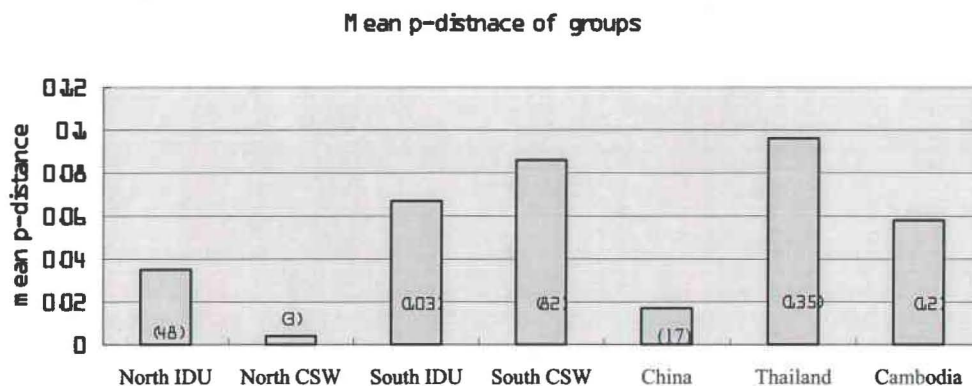


FIG. 2. Overall mean nucleotide p-distance by category. MEGA2 program was employed to calculate the mean p-distances of the following areas: (1) North Vietnam IDU, (2) North Vietnam CSW, (3) South Vietnam IDU, (4) South Vietnam CSW, (5) China, (6) Thailand, and (7) Cambodia. The numbers of the samples were in the parenthesis

Minimum-Evolution Tree: FIG. 3 was a phylogenetic tree constructed by the minimum evolution method, which could classify 4 groups evolved from 4 clusters: group I (n=219), II (n=55), III (n=30), and IV (n=98).

We sought a correlation between the nodes and the population. FIG. 4 phase A showed the correlations between the group and geographic area or risk factor: (Group I) 36 (16%) South Vietnam IDU, 55 (25%) South Vietnam CSW, 1 (0.04%) North Vietnam IDU, 2 (0.08%) China, 7 (0.3%) Cambodia, 117 Thailand (53%), and 1 Myanmar (0.04%), (Group II) 2 (4%) South

Vietnam IDU, 2 (4%) South Vietnam CSW, 33 (60%) North Vietnam IDU, 3 (5%) North Vietnam CSW, 13 (24%) China, 1 (2%) Cambodia, and 1 (2%) Thailand, (Group III) 10 (33%) South Vietnam IDU, 4 (13%) South Vietnam CSW, 1 (3%) North Vietnam IDU, 3 (10%) Cambodia, 11 (37%) Thailand, and 1 (3%) Malaysia, (Group IV) 55 (56%) South Vietnam IDU, 21 (21%) South Vietnam CSW, 13 (13%) North Vietnam, 2 (2%) China, 1 (1%) Cambodia, and 6 (6%) Thailand.

Both IDU and CSW from North Vietnam clustered mostly in group II. Group II had all three CSW and 69% (33 of 48) IDU in North Vietnam.

Both IDU and CSW from South Vietnam mostly clustered in group I and IV. CSW tend to cluster in group I (67%, 55 of 82) while IDU to group IV (53%, 55 of 103).

The sequences from Thailand clustered into group I, similar to CSW in South Vietnam (87%, 117 of 135). On the other hand, CRF01_AE sequences from China clustered in group II (77%, 13 of 17). CRF01_AE sequences from Malaysia and Myanmar clustered in group III and I, respectively. This grouping method, based on the minimum-evolution tree, successfully distinguished the genetic similarity and difference among risk factor(s) and geographic area.

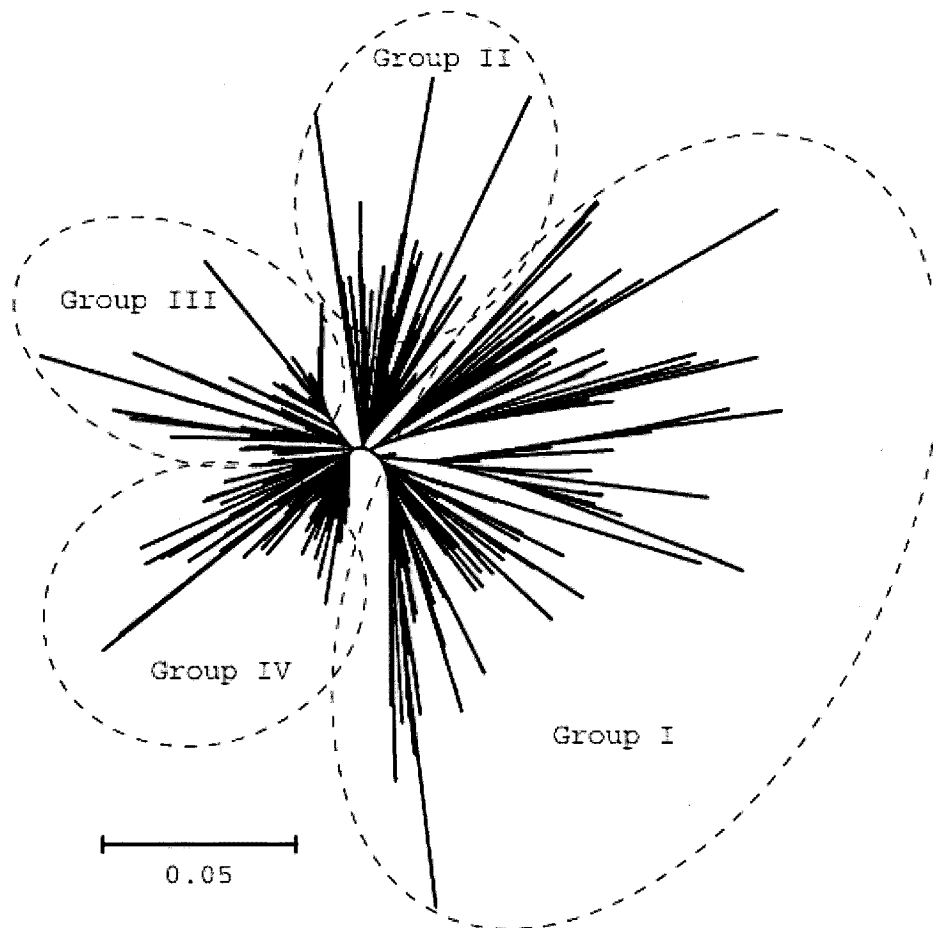


FIG. 3. A Minimum-Evolution tree was constructed by 402 CEF02_AE sequences. Four groups (Group I to IV) were defined based on four nodes.

V3 motifs and N-linked glycosylation sites: The GPGQ was predominant in this study. In TABLE 3A, 100% of CSW (3 of 3), 93% of IDU (45 of 48) from North Vietnam, 79% of CSW (65 of 82), and 83% of IDU (86 of 103) from South Vietnam had GPGQ in the V3 loop, position 312 to 315. While all of China (n=17) and Cambodia (n=12) had GPGQ, Thailand had GPGQ (43%, 59 of 135), GPGR (30%, 41 of 135), GPGH (11%, 16 of 135), GPGK (2%, 4 of 135), GGGQ (2%, 4 of 135), GQGR (n=3), GLGK (n=1), RPGQ (n=1), GQGQ (n=1), GLGQ (n=1), GPAR (n=1), GPGY (n=1), GPRR (n=1), and EPGR (n=1). All subtype B sequences from South Vietnam had GPGR.

The distribution of glycosylation sites was generally high in CRF01_AE sequences in Vietnam. TABLE 3B showed the average distributions of all six glycosylation sites: 47(98%) of IDU and 3(100%) of CSW from North Vietnam, 97(94%) of IDU, 76(93%) of CSW from South Vietnam, 17(100%) of China, 115(86%) of Thailand, 12(100%) of Cambodia, and 3(75%) of subtype B in South Vietnam.

Viral Epidemiological Signature Pattern Analysis (VESPA): We executed VESPA program to identify the signature patterns of risk groups. We have previously reported that the methionine substitution at position 309 (HXB2 numbering system, I309M) is dominant in Vietnam (48), and the valine substitution at position 343 (K343V) is common in North Vietnam (24, 25). VESPA successfully identified these two signature patterns and more: cytosine substitution at nucleotide position 7269 (T7269C).

Table 4 summarized these substitutions as follows: the methionine substitution (I309M) was frequently observed in South Vietnam: 60% IDU and 37% CSW. The strains from North Vietnam predominantly have the valine substitutions (K343V), 64% and 100% among IDU and CSW, respectively.

TABLE 3A: The tetrapeptide motif of the V3 loop.

	North Vietnam		South Vietnam		Thailand	Cambodia	China	Subtype B
	CSW	IDU	CSW	IDU				
GPGR	3(100%)	45(93%)	65(79%)	86(83%)	59(43%)	12(100%)	17(100%)	
GPGR		2(4%)	10(12%)	7(6%)	41(30%)			4(100%)
GPGR			2(2%)	2(1%)	16(11%)			
GPGR			5(6%)	8(7%)	4(2%)			
GGGR					4(2%)			
Others		1(2%)*			15(11%)**			
Total	3	48	82	103	135	12	17	4

* RPRQ was in IDU from North Vietnam

** The other patterns in Thailand were GQGR (n=3), GLGK (n=1), RPRQ (n=1), GQGR (n=1), GLGQ (n=1), GPAR (n=1), GPGY (n=1), GPRR (n=1), and EPGR (n=1)

TABLE 3B: Presence of glycosylation site* at position 280, 289, 295, 301, 334, and 356

Glycosylation	North Vietnam		South Vietnam		Thailand	Cambodia	China	Subtype B
	CSW	IDU	CSW	IDU				
280	3(100%)	46(95%)	82(100%)	103(100%)	132(98%)	12(100%)	17(100%)	3(75%)
289	3(100%)	47(98%)	82(100%)	102(99%)	127(94%)	12(100%)	17(100%)	3(75%)
295	3(100%)	48(100%)	69(84%)	96(93%)	130(96%)	12(100%)	16(94%)	2(50%)
301	3(100%)	47(98%)	72(88%)	90(87%)	63(46%)	10(83%)	17(100%)	4(100%)
334	3(100%)	48(100%)	76(93%)	96(93%)	129(96%)	12(100%)	15(88%)	1(25%)
356	3(100%)	46(95%)	75(91%)	92(89%)	113(83%)	12(100%)	17(100%)	4(100%)
Average	3(100%)	47(98%)	76(93%)	97(94%)	115(86%)	12(100%)	17(100%)	3(75%)

*Glycosylation sites typically have the motif of N-X-S or T. In our sequenced range, there were 5 possible glycosylation sites.

The signature substitutions were not only based on the peptide change, but also on nucleic acid substitution. The T7269C substitution was common in all CRF01_AE sequences in Vietnam and China (TABLE 4). The prevalence was relatively similar in CRF01_AE strains from CSW in South Vietnam and from sequences from Cambodia and was only 1% among sequence from Thailand.

TABLE 4. Viral Epidemiological Signature Pattern

	n	T7269C*	I309M**	K343V***
North Vietnam				
CSW	3	3(100%)	0	3(100%)
IDU	48	45(93%)	10(20%)	31(64%)
South Vietnam				
CSW	82	39(47%)	31(37%)	1(1%)
IDU	103	86(83%)	62(60%)	2(1%)
Thailand	135	1(1%)	13(9%)	0
China	17	15(88%)	2(11%)	15(88%)
Cambodia	12	4(33%)	2(16%)	0
Myanmar	1	0	0	0
Malaysia	1	0	0	0

VESPA identified three distinctive substitutions. Bold letter means predominant in the specific population.

* A nucleotide substitution at position 7269: the thymine as a consensus base is substituted to cytosine.

** A peptide substitution at position 309: isoleucine as a consensus is substituted to methionine.

*** A peptide substitution at position 343: lysine as a consensus was substituted to valine.

These results indicated that CRF01_AE strains circulating in Southeast Asia could be classified into 4 groups according to three substitutions. TABLE 5 showed the definition of groups: ET-, ECM, ECV, and EC-, where (1) the ET- strains were characteristic of having thymine substitution at position 7269 and no methionine and no valine substitutions at positions 309 and 343, (2) the ECM strains had the trait of methionine substitution at position 309 and cytosine at position 7269 but did not have the valine substitution at nucleotide position 343, (3) the ECV type had distinctive valine substitution at position 343, cytosine at position 7269 and lacked the

methionine substitution at position 309, and (4) the EC- strains seemed intermediate between ET- and ECM, having cytosine at 7269 but no methionine or valine at 309 and 343.

TABLE 5. Definition of four variant subtypes by VESPA method

Group by VESPA	7269	309	343	Corresponding Group	Predominant Area
ET-	T	NOT "M"	NOT "V"	Group I	Thailand, Cambodia, CSW in South Vietnam
ECM	C	M	NOT "V"	Group IV	IDU and CSW in South Vietnam
ECV	C	NOT "M"	V	Group II	North Vietnam and China
EC-	C	NOT "M"	NOT "V"	N/A	IDU and CSW in South Vietnam and IDU in North Vietnam

Four groups were defined by three distinctive substitutions.
 ET- lacked all three substitutions including T7269C, I309M, and K343V.
 ECM had T7269C and I309M but no K343V.
 ECV had T7269C and K343V but no I309M.
 EC- had T7269C but absent of I309M and K343V.

There was a correlation among the four types and geographic area or risk factor. FIG. 4

phase B showed:

- (a) The ET- strains were found mostly among CSW in South Vietnam (43%, 35 of 82), Cambodia (67%, 8 of 12), Thailand (90%, 121 of 135), Malaysia (100%, 1 of 1), and Myanmar (100%, 1 of 1).
- (b) The ECM strains were predominant among IDU in South Vietnam (56%, 58 of 103). Although the prevalence was lower than IDU, CSW in South Vietnam had 28% (23 of 82) ECM type sequences.
- (c) The ECV type was mostly found among both CSW (100%, 3 of 3) and IDU (63%, 30 of 48) strains from North and also in China (87%, 13 of 15).

- (d) The EC- strains were predominantly seen among IDU (25%, 26 of 103) and CSW (18%, 15 of 82) in South Vietnam.

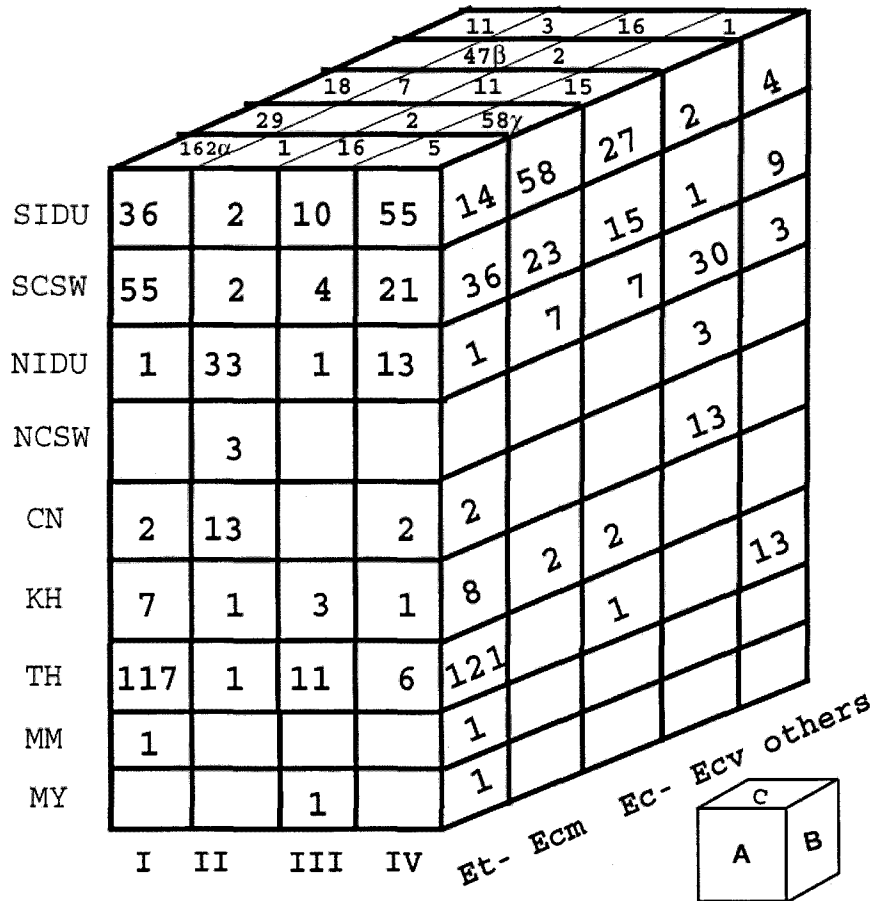


FIG.4. Distribution of the variant subtypes grouped by minimum evolution and VESPA methods in areas and risk groups: South Vietnam IDU (SIDU), South Vietnam CSW (SCSW), North Vietnam IDU (NIDU), North Vietnam CSW (NCSW), China (CN), Cambodia (KH), Thailand (TH), Myanmar (MM), and Malaysia (MY).

The phase A shows the relationship between tree based classification and risk groups in the certain areas. The phase B shows the correlations between VESPA based classification and risk groups in the areas.

The phase C shows the correlations between tree and VESPA based classification.

- α group I was correlated with ET- ($p < 0.05$)
- β group II was correlated with ECV ($p < 0.05$)
- χ group IV was correlated with ECM ($p < 0.05$)

Synonymous and Non-synonymous distance (dN/dS): After classifying the sequences with glycan at 301 and without it, we calculated the dN/dS value in strains obtained from IDU (n=47) and CSW (n=3) from North Vietnam, and IDU (n=90) and CSW (n=72) from South Vietnam with glycan. In South Vietnam, the numbers of IDU and CSW without glycan were 13 and 10 respectively. The North Vietnam population size was too small to have statistical significance.

TABLE 6 showed dN/dS value in IDU from North Vietnam, IDU and CSW from South Vietnam: IDU from North Vietnam had the dN/dS ratio 1.429 demonstrating selective pressure. IDU and CSW in South Vietnam with glycosylation site had 1.645 and 1.951 respectively. A previous study in Singapore reported that a dN/dS value in the V3 loop with glycan was 1.001 while 2.1712 for without glycan (22). Our data also showed the higher dN/dS values in with glycan: 1.645 with glycan vs. 2.173 without glycan among IDU in South Vietnam and 1.951 with glycan vs. 2.273 without glycan among CSW in South Vietnam.

For further analysis, we calculated the dN/dS for non-hypermutated NSI sequences in group I, and IV. The dN/dS of IDU (n=16) and CSW (n=28) in South Vietnam, belonging to group I, was 2.200 and, respectively. The dN/dS ratio of South Vietnam IDU (n=44) and CSW (n=17) sequences belonging to group IV was 0.514 and 0.651, respectively. These results did not support our hypothesis: the dN/dS ratio was different between IDU and CSW.

TABLE 6. The dN/dS ratio

Risk	n	dN	dS	dN/dS
North Vietnam*				
IDU	47	0.030	0.021	1.429
CSW	3	0.000	0.000	0.000
South Vietnam IDU				
w/ glycosylation**	90	0.051	0.031	1.645
w/o glycosylation	13	0.176	0.081	2.173
South Vietnam CSW				
w/ glycosylation	72	0.080	0.041	1.951
w/o glycosylation	10	0.100	0.044	2.273
South Vietnam in Group I***				
IDU	16	0.066	0.030	2.200
CSW	28	0.076	0.050	1.520
South Vietnam in Group IV***				
IDU	44	0.019	0.037	0.514
CSW	17	0.041	0.063	0.651

* In North Vietnam, 1 sample which does not have glycosylation site at 301 was removed.

** w/glycosylation has a glycosylation site at 301, and w/o glycosylation does not have.

*** SI and hypermutated sequences were removed from Group I and IV. All samples have glycosylation site at 301

The dN/ds value (Non-synonymous distance/Synonymous distance) in the V3 loop was calculated by MEGA2. The model was Modified Nei-Gojobori (jukes-Cantor method).

Evolutionary Rate: TipDate program successfully estimated the absolute rate of molecular evolution. TABLE 7 showed the evolutionary rate of each group: group I had 0.0059 (0.0029-0.0082), CSW of group I had 0.0048 (0.0024-0.0074), group II had 0.0000 (0.0000-0.0085), IDU of group II also had 0.0000 (0.0000-0.0013), group III had 0.0022 (0.0000-0.0068), group IV had 0.0035 (0.0012-0.0057), and CSW of group IV had 0.0028.

TABLE 7: The absolute rate of molecular evolution

Populatio	Risk	n	Rate	Range
Group I	All*	44	0.0059	(0.0029-0.0082)
	CSW	28	0.0048	(0.0024-0.0074)
Group II	All	38	0.0000	(0.0000-0.0085)
	IDU	13	0.0000	(0.0000-0.0013)
Group III	All	33	0.0022	(0.0000-0.0068)
Group IV	All	73	0.0035	(0.0012-0.0057)
	IDU	56	0.0028	(0.0017-0.0066)

*All are both IDU and CSW

The estimation was done by FastDNAm1 and TipDate program. Rate is an absolute rate of molecular evolution. We deleted IDU from group I and CSW from group II and group IV to specify the correlation by risk factors.

Chapter 4. Discussion

Education and vaccine are two essential tools to control infectious diseases. Education is the most important tool to prevent HIV infection until vaccine is developed. Thailand successfully slowed HIV infection rate with an intensive awareness campaign and educational program. A “100% condom program” supported by government and non-government organization encouraged condom use with each encounter. They focused on CSW and their customers, and condom use significantly increased. The cases of STD and HIV decreased as the number of CSW and their customers reduced (9, 12, 16, 19, 66-68, 70). HIV prevalence among all adults fell from 4% in 1999 to 2.6% in 2002 (13).

An effective vaccine has not been developed yet; however, the phase III trial for AIDSVAX B/E vaccine trial among IDU from Thailand was started in May 1999, and the results will be available in late 2003 (17). AIDSVAX B/E is a bivalent vaccine, consisting of a preparation of recombinant gp120 from MN isolate of subtype B and CM244 of CRF01_AE. It is noteworthy to mention that CM244 is a non-syncytium inducing strain while MN isolate is a syncytium inducing strain.

In this study, we first employed the minimum evolution method to classify the variant subtypes because it was quite possible for multiple risk groups to exchange viruses in their populations; as a result, one population might have different types of strains. Although the maximum likelihood method generated evolutionarily more reliable results than other phylogenetic methods, it required an enormous amount of time and computational power. Therefore, we chose the minimum evolution method under the MEGA2 program and classified the samples into group I, II, III, and IV based on the four nodes. This method was fast and easy, and MEGA2 was user friendly and available on public domain website.

We suggested the VESPA based classification since it is user friendly and required minimum of only three sequences. Additionally, the VESPA method could classify more clearly the nodes than tree-based method. The VESPA method could classify into ET-, ECM, EC-, and ECV. These two classification methods were highly correlated. FIG. 4 phase C showed that: (α) group I was correlated with ET- ($p < 0.05$); (162 ET-, 29 ECM, 18 EC-, and 11 others), (β) group II was correlated with ECV ($p < 0.05$); (1 ET-, 7 EC-, 47 ECV, and 3 others) and (χ) group IV was correlated with ECM ($p < 0.05$); (5 ET-, 58 ECM, 15 EC-, and 1 other). Group III was not correlated with the groups classified by the VESPA method. Group III possibly had intermediate HIV strains.

The correlations was masked by other study populations, therefore we analyzed only Vietnam samples. We had 236 CRF01_AE Vietnam sequences which were classified in to group I (92 sequences), group II (40 sequences), group III (15 sequences), and group IV (89 sequences). Group I consisted of 45 ET-, 28 ECM, 17 EC-, and 2 others ($p=0.4$). It suggested that group I was correlated to ET- in Thailand; however, there was no statistical significance in Vietnam.

On the contrary, Group II and IV were correlated with their respective VESPA classification (Group II, with ECV ($p<0.05$) and Group IV with ECM ($p<0.05$).

The genotyping classification is meaningful epidemiologically and phenotypically. The epidemiological data gives the information about the infectious routes and patients' behavior and is useful for educational program. The phenotypic information can reveal the difference or similarity of biological properties and is useful for vaccine development.

Epidemiological Information: Previous reports suggested that HIV was introduced from Thailand to Vietnam via CSW (42, 48). Our results supported this hypothesis based on mean p-distance and distribution of SI viruses and hypermutation and we suggest that the first outbreak of HIV occurred among CSW in South Vietnam followed by and IDU in South Vietnam and IDU in North Vietnam.

CSW in South Vietnam had the highest p-distance value, having more hypermutated and SI viruses than other categories in Vietnam, which indicated high diversity. The minimum evolution and VESPA methods revealed that CSW in South Vietnam and Thailand shared the same type of strain called Et- or group I.

Collectively, our data suggest that these strains were introduced from Thailand in to CSW in South Vietnam at the early stage of the outbreak. This strain of HIV was also observed in Cambodia; supporting the epidemiological reports that majority of the risk group was CSW in this country.

Next, HIV appeared among IDU in South Vietnam. ECM or group IV was the main type in this group. Although CSW in South Vietnam composed mainly of Et-, it had ECM as well, and vice versa for IDU in South Vietnam. This suggests a possible bridge between the two risk groups. Some patients overlapped in both risk groups and became spreaders.

HIV was introduced to North Vietnam independently and more recently than South Vietnam. While there were genetic differences between IDU and CSW in South Vietnam, in North Vietnam IDU were similar to CSW. The low genetic diversity and few SI or hypermutated sequences supported our conclusion. Also the viruses in North Vietnam were ECV which were genetically different from viruses mainly found in South Vietnam.

While the educational campaign in Thailand is focused on CSW, in Vietnam the campaign is focused on the control of HIV infection among IDU rather than CSW, and one of the most prominent campaigns is a syringe-exchange program that allows IDU to use a clean syringe (14).

To minimize the cost and to maximize the efficacy of HIV control programs, it is necessary to know the vulnerable risk populations. We measured the median age of each risk population and found that certain age groups existed. As expected, CSW were young women; the median age was 24 yrs (n=72, range 15 – 64). The median age of IDU men in South Vietnam was 40 years (n=77, range 15 – 56) while two IDU women in South Vietnam were 30 and 38 years old, older than the median age of CSW. These two IDU women belonged to group IV in the minimum evolution method and ECM and EC- in the VESPA method. These groups were seen predominantly among IDU in South Vietnam.

IDU in North Vietnam were on the other hand, younger than the IDU in South Vietnam: median age 26 years (n=45, range 17 – 38 years). The median age of STD patients in South Vietnam differed by men and women: 34 years (range 19 – 47 years) for men and 23 years (range 20 – 27) for women.

In summary, HIV infected IDU in South Vietnam were middle aged, irrespective of the gender on the other hand CSW in South Vietnam were young. In North Vietnam, both IDU and CSW were young. Therefore, the educational program has to target among men over 30 years old in IDU population in South Vietnam. On the contrary, CSW in South Vietnam and IDU and CSW in North Vietnam need HIV/AIDS control program for young generation: less than 30 years old.

Phenotypical Information: Genotype often impacts their phenotype; differences among viral characteristics have to be reviewed before designing a vaccine. We will discuss four important factors in this section: (1) subtype, (2) cell tropism, (3) glycosylation site, and (4) transmission pathway.

Subtype: While subtypes B', C, or C/B' recombinants were prevalent among southeast Asian countries, such as Myanmar, China, Thailand and Singapore, the predominant subtype in Vietnam was CRF01_AE from 1995 to 1999 (96.7% were CRF01_AE, and 3.3% were subtype B). Subtype B' mainly circulates in Thailand, and C and C/B' recombinant have been seen in Southern China from the mid 1990's. Therefore, the introduction from China or Thailand did not come to the

surface, and we suggest focusing on a vaccine for CRF01_AE in Vietnam intensively, rather than subtype B', C, or C/B'.

Cell tropism: Due to immunological and other selection pressures, HIV switches its tropism in the host. During early infection, HIV utilizes the chemokine receptor CCR5 chemokine co-receptor for entrance into target cells. This type of virus is called R5, and is usually macrophage tropic and NSI. Later, it utilizes the chemokine receptor CXCR4, commonly called as X4, and is usually T cell tropic and SI virus. We classified a virus as NSI or SI by identifying substitutions with high charged amino acid such as arginine (R) or histidine (H) at six positions (301, 303, 306, 308, 315, and 322). This criterion was not supported statistically, but a trend existed since the viruses were collected from AIDS patients.

One of the marker positions was at the end of tetrapeptide crown motif in the V3 loop. The predominant V3 tetrapeptide crown motif was GPGQ in CRF01_AE while subtype B displayed an extensive variation, and higher charged types including GPGR and GPGH were highly observed (79). Therefore, CRF01_AE mostly showed NSI traits, and AIDSVAX B/E vaccine also has a backbone of NSI strain. In other words, the vaccine was not designed for SI viral strains of CRF01_AE, and studying SI viruses is urgently warranted.

The history of HIV endemic and SI distribution were correlated in Vietnam. For example, in IDU and CSW from North Vietnam, where HIV was relatively recently introduced, the HIV strains were mostly NSI distribution (94%). Whereas in Thailand, where the HIV epidemic has a long history had 65% SI viral strains. The SI strain distribution in CSW and IDU in South Vietnam was 31% and 28%, respectively, and the viral strains are been circulating in this high-risk population for relatively longer than North Vietnam but shorter than Thailand.

Minimum evolution tree also supported our thesis that longer the virus circulates in a population more SI strains are generated. Based on the minimum evolution method, the group I sequences, most of whom were from South Vietnam and Thailand had more SI viral strains (45%) than any other groups, 2% in group II, 6% in group III, and 15% in group IV.

As previously reported, lack of glycosylation site 280, 289, 295, 301, and 334 is correlated to SI viruses (22). Glycans have potential to mask epitopes, and viruses lacking glycosylation site would be more susceptible to immune surveillance and possibly evade from the host immunity. Similar to SI genotype, lack of glycans was seen in Thailand (17% average) more frequently than IDU (2% average) and CSW (zero) in North Vietnam, IDU (6% average) and CSW (6% average) in South Vietnam, China (zero), or Cambodia (zero).

These results supported our epidemiologic conclusions: HIV appeared in Thailand, followed by CSW and IDU in South Vietnam, and later among CSW and IDU in North Vietnam chronologically. In conclusion, as SI variants are seen more often in Vietnam, it is important to study the efficacy of vaccine against SI strains.

Viral differences between transmission pathway (IDU and CSW): A previous study reported lower HIV genetic diversity in IDU than among individuals infected via heterosexual transmission (59). Several studies in IDU demonstrated genetic distance less than 0.02 (8, 15, 35-37, 50, 51, 59) while over 0.08 among patients infected via heterosexual transmission (11, 51). The plausible reasons were: (1) viruses circulating among patients infected by heterosexual transmission were selected through mucosal barrier; on the other hand, IDU did not give any selective pressure on viruses, (2) a single introduction of HIV with higher infection rate in IDU cause an intensive founder effect, or (3) simply viruses among IDU was younger than viruses among sexually transmitted patients.

Our study also showed 0.086 genetic distances in CSW (n=82) in South Vietnam. However, the mean p-distance of IDU (n=103) in South Vietnam was 0.067, higher than the genetic distance reported in previous IDU studies. IDU (n=48) in North Vietnam showed low

diversity, 0.035. CSW (n=3) in North Vietnam was expected to be higher than IDU according to the previous reports. However, the sequences were closely related to each other, and the genetic distance was 0.004. Recent reports have suggested use of injection drugs among CSW in North Vietnam, further suggesting that the viruses in CSW was probably acquired via IDU rather than via heterosexual sex. Moreover, VESPA analysis further confirmed this thesis, since the amino acid and nucleic acid substitutions in CSW from North Vietnam was very similar to North Vietnam IDU as well as HIV strains found in South China.

To prove the hypothesis that viruses in CSW differed from ones in IDU due to different infectious pathways, we applied the dN/dS analysis and evaluated the selective pressure. While IDU and CSW in group I showed different dN/dS, the results were not consistent among other groups.

For further analysis, we estimated the evolutionary rate in groups I, II, III, and IV. Group II and IV represented viruses circulating in IDU in North and South Vietnam, respectively. On the other hand, group I represented ones among CSW. Group I showed higher evolutionary rate, (0.0059) than group II (0.000) or group IV (0.0035). However, these representative groups contained both IDU and CSW; therefore, we deleted IDU from group I and CSW from group II and IV to study the differences in genetic differences based on risk factor for acquisition of HIV.

Group I consisted of only CSW (n=28) had the evolutionary rate, 0.0048. Group II (n=33) and group IV consisted of only IDU had the evolutionary rate, 0.000 and 0.0028, respectively. It suggested that the evolutionary speed of IDU was slower than CSW; however, there might be a gap between collection year and infected year.

Although the dN/dS value did not show any significant correlations, the evolutionary rate was higher in CSW than in IDU.

Viral types circulating in Vietnam had unique viral epidemiological sequences, which will in future allow us to track the routes of infection and as well as trafficking of HIV in southeast Asia. The viral traits have to be considered as designing the educational program and vaccine program.

References

1. 2001. Geography Map of HIV Database.Los Alamos National Laboratory.
2. 2002.More than 51,500 HIV-Positive in Vietnam?Agence France Presse.
3. Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman 1990. Basic local alignment search tool?J Mol Biol.215:403-10.
4. Anderson, J. P., A. G. Rodrigo, G. H. Learn, A. Madan, C. Delahunty, M. Coon, M. Girard, S. Osmanov, L. Hood, and J. I. Mullins 2000. Testing the hypothesis of a recombinant origin of human immunodeficiency virus type 1 subtype E?J Virol.74:10752-65.
5. Aquino-De Jesus, M. J., C. Anders, G. Miller, J. W. Sleasman, M. M. Goodenow, and W. A. Andiman 2000. Genetically and epidemiologically related "non-syncytium-inducing" isolates of HIV-1 display heterogeneous growth patterns in macrophages?J Med Virol.61:171-80.
6. Beatrice H. Hahn, G. M. S., Kevin M. De, Cock, and Paul M. Sharp 2000. AIDS as a Zoonosis: Scientific and Public Health Implications?Science.28.

7. Bette, T., B. T. Korber, Foley, C. L. Kuiken, S. K. Pillai, and J. G. Sodroski?2001?posting date.Numbering Positions in HIV Relative to HXB2CG.Los Alamos HIV Sequence Database.[Online.]
8. Bobkov, A., E. Kazennova, L. Selimova, M. Bobkova, T. Khanina, N. Ladnaya, A. Kravchenko, V. Pokrovsky, R. Cheingsong-Popov, and J. Weber 1998. A sudden epidemic of HIV type 1 among injecting drug users in the former Soviet Union: identification of subtype A, subtype B, and novel gagA/envB recombinants?AIDS Res Hum Retroviruses.14:669-76.
9. Boyles, S. 1995. Decline in HIV reported among young males in Thailand?AIDS Wkly:15.
10. Brown, T. M., K. E. Robbins, M. Sinniah, T. S. Saraswathy, V. Lee, L. S. Hooi, B. Vijayamalar, C. C. Luo, C. Y. Ou, J. Rapier, G. Schochetman, and M. L. Kalish 1996. HIV type 1 subtypes in Malaysia include B, C, and E?AIDS Res Hum Retroviruses.12:1655-7.
11. Carr, J. K., T. Laukkanen, M. O. Salminen, J. Albert, A. Alaeus, B. Kim, E. Sanders-Buell, D. L. Birx, and F. E. McCutchan 1999. Characterization of subtype A HIV-1 from Africa by full genome sequencing?Aids.13:1819-26.

12. Celentano, D. D., K. E. Nelson, C. M. Lyles, C. Beyrer, S. Eiumtrakul, V. F. Go, S. Kuntolbutra, and C. Khamboonruang 1998. Decreasing incidence of HIV and sexually transmitted diseases in young Thai men: evidence for success of the HIV/AIDS control and prevention program?Aids.12:F29-36.
13. Cohen, J. 2003. Asia--the next frontier for HIV/AIDS. Two hard-hit countries offer rare success stories: Thailand and Cambodia?Science.301:1658-62.
14. Cohen, J. 2003. HIV and heroin: a deadly international affair: Vietnam?Science.301:1657-8.
15. Ferdats, A., V. Konicheva, I. Dievberna, E. Lilja, and J. Albert 1999. An HIV type 1 subtype A outbreak among injecting drug users in Latvia?AIDS Res Hum Retroviruses.15:1487-90.
16. Ford, N., and S. Koetsawang 1999. A pragmatic intervention to promote condom use by female sex workers in Thailand?Bull World Health Organ.77:888-94.
17. Francis, D. P., T. Gregory, M. J. McElrath, R. B. Belshe, G. J. Gorse, S. Migasena, D. Kitayaporn, P. Pitisuttitham, T. Matthews, D. H. Schwartz, and P. W. Berman 1998. Advancing AIDSVAX to phase 3. Safety, immunogenicity, and plans for phase 3?AIDS Res Hum Retroviruses.14:S325-31.

18. Gibas, C., and P. Jambeck 2001. Bioinformatics Computer Skills.O'Reilly.
19. Hanenberg, R., and W. Rojanapithayakorn 1996. Prevention as policy: how Thailand reduced STD and HIV transmission?Aidscriptions.3:24-7.
20. Hoshino, H., R. A. Weiss, P. Clapham, M. Miwa, I. Miyoshi, M. Yoshida, and T. Sugimura 1984. Pseudotype viruses bearing envelope antigens of Japanese isolates of human T-cell leukemia viruses type I?Princess Takamatsu Symp.15:159-64.
21. Ichimura, H., S. C. Kliks, S. Visrutaratna, C. Y. Ou, M. L. Kalish, and J. A. Levy 1994. Biological, serological, and genetic characterization of HIV-1 subtype E isolates from northern Thailand?AIDS Res Hum Retroviruses.10:263-9.
22. Kalish, M. L., B. T. Korber, S. Pillai, K. E. Robbins, Y. S. Leo, A. Saekhou, I. Verghese, P. Gerrish, C. L. Goh, D. Lupo, B. H. Tan, T. M. Brown, and R. Chan 2002. The sequential introduction of HIV-1 subtype B and CRF01AE in Singapore by sexual transmission: accelerated V3 region evolution in a subpopulation of Asian CRF01 viruses?Virology.304:311-29.
23. Kalish, M. L., C. C. Luo, B. G. Weniger, K. Limpakarnjanarat, N. Young, C. Y. Ou, and G. Schochetman 1994. Early HIV type 1 strains in Thailand were not

- responsible for the current epidemic?AIDS Res Hum Retroviruses.10:1573-5.
24. Kato, K., S. Kusagawa, K. Motomura, R. Yang, T. Shiino, K. Nohtomi, H. Sato, K. Shibamura, T. H. Nguyen, K. C. Pham, H. T. Pham, C. T. Duong, D. T. Bui, T. L. Hoang, Y. Nagai, and Y. Takebe 2001. Closely related HIV-1 CRF01_AE variant among injecting drug users in northern Vietnam: evidence of HIV spread across the Vietnam-China border?AIDS Res Hum Retroviruses.17:113-23.
25. Kato, K., T. Shiino, S. Kusagawa, H. Sato, K. Nohtomi, K. Shibamura, T. H. Nguyen, K. C. Pham, X. L. Truong, H. A. Mai, T. L. Hoang, G. Bunyaraksyotin, Y. Fukushima, M. Honda, C. Wasi, S. Yamazaki, Y. Nagai, and Y. Takebe 1999. Genetic similarity of HIV type 1 subtype E in a recent outbreak among injecting drug users in northern Vietnam to strains in Guangxi Province of southern China?AIDS Res Hum Retroviruses.15:1157-68.
26. Korber, B., M. Muldoon, J. Theiler, F. Gao, R. Gupta, A. Lapedes, B. H. Hahn, S. Wolinsky, and T. Bhattacharya 2000. Timing the ancestor of the HIV-1 pandemic strains?Science.288:1789-96.
27. Korber, B., and G. Myers 1992. Signature pattern analysis: a method for assessing viral sequence relatedness?AIDS Res Hum Retroviruses.8:1549-60.

28. Korber, B. T., S. Osmanov, J. Esparza, and G. Myers 1994. The World Health Organization Global Programme on AIDS proposal for standardization of HIV sequence nomenclature. WHO Network for HIV Isolation and Characterization?AIDS Res Hum Retroviruses.10:1355-8.
29. Kumar, S., K. Tamura, and M. Nei 1994. MEGA: Molecular Evolutionary Genetics Analysis software for microcomputers?Comput Appl Biosci.10:189-91.
30. Kusagawa, S., H. Sato, K. Kato, K. Nohtomi, T. Shiino, C. Samrith, H. B. Leng, T. Phalla, M. B. Heng, and Y. Takebe 1999. HIV type 1 env subtype E in Cambodia?AIDS Res Hum Retroviruses.15:91-4.
31. Kusagawa, S., H. Sato, S. Watanabe, K. Nohtomi, K. Kato, T. Shino, M. Thwe, K. Y. Oo, S. Lwin, R. Mra, B. Kywe, S. Yamazaki, and Y. Takebe 1998. Genetic and serologic characterization of HIV type 1 prevailing in Myanmar (Burma)?AIDS Res Hum Retroviruses.14:1379-85.
32. Lee, J. S., J. G. Nam, E. Y. Kim, C. Kang, B. K. Koo, and H. W. Cho 2000. Introduction of HIV type 1 subtype E virus into South Korea?AIDS Res Hum Retroviruses.16:1083-7.
33. Lee, T. H., J. E. Coligan, and M. Essex 1984. Human T-cell leukemia virus

specific antigens?Princess Takamatsu Symp.15:197-203.

34. Liitsola, K., P. Holmstrom, T. Laukkanen, H. Brummer-Korvenkontio, P. Leinikki, and M. O. Salminen 2000. Analysis of HIV-1 genetic subtypes in Finland reveals good correlation between molecular and epidemiological data?Scand J Infect Dis.32:475-80.
35. Liitsola, K., I. Tashkinova, T. Laukkanen, G. Korovina, T. Smolskaja, O. Momot, N. Mashkilleyson, S. Chaplinskas, H. Brummer-Korvenkontio, J. Vanhatalo, P. Leinikki, and M. O. Salminen 1998. HIV-1 genetic subtype A/B recombinant strain causing an explosive epidemic in injecting drug users in Kaliningrad?Aids.12:1907-19.
36. Lukashov, V. V., R. Huismans, A. G. Rakhmanova, Z. N. Lisitsina, N. A. Akhtyrskaya, N. N. Vlasov, O. B. Melnick, and J. Goudsmit 1999. Circulation of subtype A and gagA/envB recombinant HIV type 1 strains among injecting drug users in St. Petersburg, Russia, correlates with geographical origin of infections?AIDS Res Hum Retroviruses.15:1577-83.
37. Lukashov, V. V., E. V. Karamov, V. F. Eremin, L. P. Titov, and J. Goudsmit 1998. Extreme founder effect in an HIV type 1 subtype A epidemic among drug users in

- Svetlogorsk, Belarus?AIDS Res Hum Retroviruses.14:1299-303.
38. Machuca, R., M. Bogh, M. Salminen, J. Gerstoft, B. Kvinesdal, C. Pedersen, N. Obel, H. Nielsen, and C. Nielsen 2001. HIV-1 subtypes in Denmark?Scand J Infect Dis.33:697-701.
39. McCutchan, F. E., A. W. Artenstein, E. Sanders-Buell, M. O. Salminen, J. K. Carr, J. R. Mascola, X. F. Yu, K. E. Nelson, C. Khamboonruang, D. Schmitt, M. P. Kieny, J. G. McNeil, and D. S. Burke 1996. Diversity of the envelope glycoprotein among human immunodeficiency virus type 1 isolates of clade E from Asia and Africa?J Virol.70:3331-8.
40. McCutchan, F. E., P. A. Hegerich, T. P. Brennan, P. Phanuphak, P. Singharaj, A. Jugsudee, P. W. Berman, A. M. Gray, A. K. Fowler, and D. S. Burke 1992. Genetic variants of HIV-1 in Thailand?AIDS Res Hum Retroviruses.8:1887-95.
41. McCutchan, F. E., K. Viputtigul, M. S. de Souza, J. K. Carr, L. E. Markowitz, P. Buapunth, J. G. McNeil, M. L. Robb, S. Nitayaphan, D. L. Birx, and A. E. Brown 2000. Diversity of envelope glycoprotein from human immunodeficiency virus type 1 of recent seroconverters in Thailand?AIDS Res Hum Retroviruses.16:801-5.

42. Menu, E., T. X. Truong, M. E. Lafon, T. H. Nguyen, M. C. Muller-Trutwin, T. T. Nguyen, A. Deslandres, G. Chaouat, Q. T. Duong, B. K. Ha, H. J. Fleury, and F. Barre-Sinoussi 1996. HIV type 1 Thai subtype E is predominant in South Vietnam?AIDS Res Hum Retroviruses.12:629-33.
43. Motomura, K., S. Kusagawa, K. Kato, K. Nohtomi, H. H. Lwin, K. M. Tun, M. Thwe, K. Y. Oo, S. Lwin, O. Kyaw, M. Zaw, Y. Nagai, and Y. Takebe 2000. Emergence of new forms of human immunodeficiency virus type 1 intersubtype recombinants in central Myanmar?AIDS Res Hum Retroviruses.16:1831-43.
44. Mullis, K., F. Faloona, S. Scharf, R. Saiki, G. Horn, and H. Erlich 1986. Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction?Cold Spring Harb Symp Quant Biol.51:263-73.
45. Murphy, E., B. Korber, M. C. Georges-Courbot, B. You, A. Pinter, D. Cook, M. P. Kieny, A. Georges, C. Mathiot, F. Barre-Sinoussi, and et al. 1993. Diversity of V3 region sequences of human immunodeficiency viruses type 1 from the central African Republic?AIDS Res Hum Retroviruses.9:997-1006.
46. Nei, M., and T. Gojobori 1986. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions?Mol Biol

Evol.3:418-26.

47. Nerurkar, V. R., H. T. Nguyen, W. M. Dashwood, P. R. Hoffmann, C. Yin, D. M. Morens, A. H. Kaplan, R. Detels, and R. Yanagihara 1996. HIV type 1 subtype E in commercial sex workers and injection drug users in southern Vietnam?AIDS Res Hum Retroviruses.12:841-3.
48. Nerurkar, V. R., H. T. Nguyen, C. L. Woodward, P. R. Hoffmann, W. M. Dashwood, H. T. Long, D. M. Morens, R. Detels, and R. Yanagihara 1997. Sequence and phylogenetic analyses of HIV-1 infection in Vietnam: subtype E in commercial sex workers (CSW) and injection drug users (IDU)?Cell Mol Biol (Noisy-le-grand).43:959-68.
49. Nguyen, T. H., B. Masquelier, V. H. Pham, T. V. Nguyen, M. E. Lafon, T. X. Truong, H. C. Nguyen, F. B. Sinoussi, and H. J. Fleury 1999. Further characterization of HIV-1 isolates from Ho Chi Minh City, Vietnam?J Acquir Immune Defic Syndr Hum Retrovirol.20:93-5.
50. Novitsky, V. A., M. A. Montano, and M. Essex 1998. Molecular epidemiology of an HIV-1 subtype A subcluster among injection drug users in the Southern Ukraine?AIDS Res Hum Retroviruses.14:1079-85.

51. Novitsky, V. A., M. A. Montano, M. F. McLane, B. Renjifo, F. Vannberg, B. T. Foley, T. P. Ndung'u, M. Rahman, M. J. Makhema, R. Marlink, and M. Essex 1999. Molecular cloning and phylogenetic analysis of human immunodeficiency virus type 1 subtype C: a set of 23 full-length clones from Botswana?J Virol.73:4427-32.
52. Olsen, G. J., H. Matsuda, R. Hagstrom, and R. Overbeek 1994. fastDNAmL: a tool for construction of phylogenetic trees of DNA sequences using maximum likelihood?Comput Appl Biosci.10:41-8.
53. Oroszlan, S., T. D. Copeland, N. R. Rice, G. W. Smythers, W. P. Tsai, Y. Yoshinaka, and K. Shimotohno 1984. Structural and antigenic characterization of the proteins of human T-cell leukemia viruses and their relationships to the gene products of other retroviruses?Princess Takamatsu Symp.15:147-57.
54. Ou, C. Y., Y. Takebe, C. C. Luo, M. Kalish, W. Auwanit, C. Bandea, N. de la Torre, J. L. Moore, G. Schochetman, S. Yamazaki, and et al. 1992. Wide distribution of two subtypes of HIV-1 in Thailand?AIDS Res Hum Retroviruses.8:1471-2.
55. Ou, C. Y., Y. Takebe, B. G. Weniger, C. C. Luo, M. L. Kalish, W. Auwanit, S. Yamazaki, H. D. Gayle, N. L. Young, and G. Schochetman 1993. Independent

- introduction of two major HIV-1 genotypes into distinct high-risk populations in Thailand?Lancet.341:1171-4.
56. Papa, A., E. Papadimitriou, A. Papoutsi, V. Kiosses, and A. Antoniadis 2002. HIV-1 subtypes and circulating recombinant forms (CRFs) in Northern Greece?Virus Res.85:85-93.
57. Perelson, A. S., A. U. Neumann, M. Markowitz, J. M. Leonard, and D. D. Ho 1996. HIV-1 dynamics in vivo: virion clearance rate, infected cell life-span, and viral generation time?Science.271:1582-6.
58. Piot, P., F. A. Plummer, F. S. Mhalu, J. L. Lamboray, J. Chin, and J. M. Mann 1988. AIDS: an international perspective?Science.239:573-9.
59. Piyasirisilp, S., F. E. McCutchan, J. K. Carr, E. Sanders-Buell, W. Liu, J. Chen, R. Wagner, H. Wolf, Y. Shao, S. Lai, C. Beyrer, and X. F. Yu 2000. A recent outbreak of human immunodeficiency virus type 1 infection in southern China was initiated by two highly homogeneous, geographically separated strains, circulating recombinant form AE and a novel BC recombinant?J Virol.74:11286-95.
60. Plantier, J. C., L. Vergne, F. Damond, M. B. S, M. P.-N. E, L. Buzelay, I. Farfara,

- D. Brand, M. Peeters, F. Brun-Vezinet, E. Delaporte, and F. Barin 2002. Development and evaluation of a DNA enzyme immunoassay method for env genotyping of subtypes A through G of human immunodeficiency virus type 1 group M, with discrimination of the circulating recombinant forms CRF01_AE and CRF02_AG?J Clin Microbiol.40:1010-22.
61. Puchhammer-Stockl, E., C. Kunz, E. Faatz, P. Kasper, and F. X. Heinz 1998. Introduction of HIV-1 subtypes C, E and A into Austria?Clin Diagn Virol.9:25-8.
62. Rambaut, A. 2000. Estimating the rate of molecular evolution: incorporating non-contemporaneous sequences into maximum likelihood phylogenies?Bioinformatics.16:395-9.
63. Reinis, M., M. Bruckova, R. R. Graham, J. Vandasova, M. Stankova, and J. K. Carr 2001. Genetic subtypes of HIV type 1 viruses circulating in the Czech Republic?AIDS Res Hum Retroviruses.17:1305-10.
64. Robertson, D. L., J. P. Anderson, J. A. Bradac, J. K. Carr, B. Foley, R. K. Funkhouser, F. Gao, B. H. Hahn, M. L. Kalish, C. Kuiken, G. H. Learn, T. Leitner, F. McCutchan, S. Osmanov, M. Peeters, D. Pieniazek, M. Salminen, P. M. Sharp, S. Wolinsky, and B. Korber 2000. HIV-1 nomenclature

- proposal?Science.288:55-6.
65. Robertson, D. L., P. M. Sharp, F. E. McCutchan, and B. H. Hahn 1995.
Recombination in HIV-1?Nature.374:124-6.
66. Rojanapitayakorn, W. 1993. "100 percent" condom use seeks to slow HIV
spread?Network.13:30, 32.
67. Rojanapithayakorn, W. 1998. Promising results. Thailand?Integration:21-3.
68. Rojanapithayakorn, W., and R. Hanenberg 1996. The 100% condom program in
Thailand?Aids.10:1-7.
69. Rose, P. P., and B. T. Korber 2000. Detecting hypermutations in viral sequences
with an emphasis on G --> A hypermutation?Bioinformatics.16:400-1.
70. Rugpao, S., S. Tovanabutra, C. Beyrer, D. Nuntakuang, Y. Yutabootr, T. Vongchak,
M. A. de Boer, D. D. Celentano, and K. E. Nelson 1997. Multiple condom use in
commercial sex in Lamphun Province, Thailand: a community-generated
STD/HIV prevention strategy?Sex Transm Dis.24:546-9.
71. Rzhetsky, A., and M. Nei 1993. Theoretical foundation of the minimum-evolution
method of phylogenetic inference?Mol Biol Evol.10:1073-95.
72. Salemi, M., K. Strimmer, W. W. Hall, M. Duffy, E. Delaporte, S. Mboup, M.

- Peeters, and A. M. Vandamme 2001. Dating the common ancestor of SIVcpz and HIV-1 group M and the origin of HIV-1 subtypes using a new method to uncover clock-like molecular evolution?Faseb J.15:276-8.
73. Sanger, F., J. E. Donelson, A. R. Coulson, H. Kossel, and D. Fischer 1974. Determination of a nucleotide sequence in bacteriophage ϕ 1 DNA by primed synthesis with DNA polymerase?J Mol Biol.90:315-33.
74. Se-Thoe, S. Y., B. T. Foley, S. Y. Chan, R. V. Lin, H. M. Oh, A. E. Ling, S. K. Chew, I. Snodgrass, and J. E. Sng 1998. Analysis of sequence diversity in the C2-V3 regions of the external glycoproteins of HIV type 1 in Singapore?AIDS Res Hum Retroviruses.14:1601-4.
75. Shieh, B., C. M. Lee, Y. M. Chen, I. J. Su, and C. Li 1997. Molecular subtyping of the HIV-1 V3 loop sequences detected in HIV-1-positive patients in southern Taiwan?J Microbiol Immunol Infect.30:106-14.
76. Simon, F., I. Loussert-Ajaka, F. Damond, S. Saragosti, F. Barin, and F. Brun-Vezinet 1996. HIV type 1 diversity in northern Paris, France?AIDS Res Hum Retroviruses.12:1427-33.
77. Stanojevic, M., A. Papa, E. Papadimitriou, S. Zerjav, D. Jevtovic, D. Salemovic, T.

- Jovanovic, and A. Antoniadis 2002. HIV-1 subtypes in Yugoslavia?AIDS Res Hum Retroviruses.18:519-22.
78. Subbarao, S., K. Limpakarnjanarat, T. D. Mastro, J. Bhumisawasdi, P. Warachit, C. Jayavas, N. L. Young, C. C. Luo, N. Shaffer, M. L. Kalish, and G. Schochetman 1998. HIV type 1 in Thailand, 1994-1995: persistence of two subtypes with low genetic diversity?AIDS Res Hum Retroviruses.14:319-27.
79. Subbarao, S., S. Vanichseni, D. J. Hu, D. Kitayaporn, K. Choopanya, S. Raktham, N. L. Young, C. Wasi, R. Sutthent, C. C. Luo, A. Ramos, and T. D. Mastro 2000. Genetic characterization of incident HIV type 1 subtype E and B strains from a prospective cohort of injecting drug users in Bangkok, Thailand?AIDS Res Hum Retroviruses.16:699-707.
80. Thompson, J. D., D. G. Higgins, and T. J. Gibson 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice?Nucleic Acids Res.22:4673-80.
81. UNAIDS 2000. Report on the global HIV/AIDS epidemic.UNAIDS.
82. Ungchusak, K., S. Sriprapandh, S. Pinichpongse, P. Kunasaol, and S.

- Thanprasertsuk 1989. First national sentinel seroprevalence survey of HIV-1 infection in Thailand?Thai AIDS J.1:57-74.
83. Wang, Z., C. M. Lyles, C. Beyrer, D. D. Celentano, D. Vlahov, C. Natpratan, R. Markham, C. Khamboonruang, K. Nelson, and X. F. Yu 1998. Diversification of subtype E human immunodeficiency virus type 1 env in heterosexual seroconverters from Northern Thailand?J Infect Dis.178:1507-11.
84. Wangroongsarb, Y., B. G. Weniger, C. Wasi, A. Traisupa, P. Kunasol, W. Rojanapithayakorn, and S. Fucharoen 1985. Prevalence of HTLV-III/LAV antibody in selected populations in Thailand?Southeast Asian J Trop Med Public Health.16:517-20.
85. Weiss, H. A., and S. Hawkes 2001. An overview of the global epidemiology of HIV/AIDS?Lepr Rev.72:92-8.
86. Weniger, B. G., K. Limpakarnjanarat, K. Ungchusak, S. Thanprasertsuk, K. Choopanya, S. Vanichseni, T. Uneklabh, P. Thongcharoen, and C. Wasi 1991. The epidemiology of HIV infection and AIDS in Thailand?Aids.5:S71-85.
87. Weniger, B. G., Y. Takebe, C. Y. Ou, and S. Yamazaki 1994. The molecular epidemiology of HIV in Asia?Aids.8:S13-28.

88. Yamaguchi, Y., and T. Gojobori 1997. Evolutionary mechanisms and population dynamics of the third variable envelope region of HIV within single hosts?Proc Natl Acad Sci U S A.94:1264-9.
89. Yoshino, N., S. Naganawa, T. Nakasone, S. Imura, T. Kita, and M. Honda 1998. Vertical transmission of human immunodeficiency virus type 1 in Japan, 1989-1997: presence of two subtypes B and E with subtype E predominance. National Cooperative Study Investigators on Vertical Transmission of HIV-1?Acta Paediatr Jpn.40:503-9.
90. Yu, X. F., J. Chen, Y. Shao, C. Beyrer, and S. Lai 1998. Two subtypes of HIV-1 among injection-drug users in southern China?Lancet.351:1250.
91. Yu, X. F., J. Chen, Y. Shao, C. Beyrer, B. Liu, Z. Wang, W. Liu, J. Yang, S. Liang, R. P. Viscidi, J. Gu, G. Gurri-Glass, and S. Lai 1999. Emerging HIV infections with distinct subtypes of HIV-1 infection among injection drug users from geographically separate locations in Guangxi Province, China?J Acquir Immune Defic Syndr.22:180-8.