CHAPTER 2

STUDY WORKS

1. Introduction

Mother-to-child transmission (MTCT) is the leading source of human immunodeficiency virus (HIV) infection in children. Transmission may occur during pregnancy (*in utero*), labor and delivery (intrapartum) or after birth, through breastfeeding. The risk of MTCT can be reduced to below 2% by interventions such as antiretroviral (ARV) prophylaxis, elective cesarean delivery, and the complete avoidance of breastfeeding [396; 222; 263].

Maternal neutralizing antibodies (NAbs) can cross the placental barrier into the fetal bloodstream, reaching high levels in the fetus at the end of pregnancy and protecting the infant against infection with numerous pathogens [119; 349]. Therefore the MTCT context allows to study the potential prophylactic effects of antibodies in preventing HIV transmission. The studies of passive immunization with human neutralizing monoclonal antibodies (HuMAbs) in the neonatal macaques model have proven that antibodies could protect against MTCT of HIV-1 infection [27; 166]. However, conflicting results have been obtained concerning the role of maternal NAbs in the MTCT of HIV [357; 162; 227; 48]. The differences in the results obtained may be due to differences in methodological approaches and in the populations studied.

Several studies have demonstrated that HIV-1-infected mothers generally have a heterogeneous viral population, whereas their infected infants have a more homogeneous virus population [428; 357; 13; 301; 410]. This indicates the presence of selective pressure resulting in the selection of a limited number of maternal viral variants for establishment of a new infection in the infants. Maternal antibodies are among the selective factors potentially responsible for this genetic bottleneck. Indeed, several studies, including a previous study from our group, have shown that non-transmitting mothers are more likely to have higher levels of NAbs, consistent with a role of these antibodies in reducing MTCT of HIV [357; 227; 47; 156; 32]. Supporting this hypothesis, recent studies have clearly shown that variants from infants are frequently resistant to neutralization by autologous maternal plasma, suggesting transmission of maternal NAbs escape variants [111; 109; 410; 432]. However, infants have occasionally been found to be infected with a heterogeneous population of multiple maternal variants [357; 301; 410; 111; 223; 338].

It has been suggested that the timing of transmission is responsible for differences in the diversity of the virus population in infants, with *in utero* transmitters more likely to transmit single or multiple maternal variants and intrapartum transmitters more likely to transmit minor HIV-1 variants [111]. The existence of different selective pressures depending on the timing of transmission was not confirmed in other studies [410; 109]. We believe that analyses of the role of passively transferred antibodies should clearly separate cases of intrapartum transmission from *in utero* transmission: only in cases of intrapartum transmission can we be sure that exposure to the virus occurs in the presence of optimal levels of IgGs.

Understanding the role of NAbs and the molecular characteristics of the HIV-1 variants involved in MTCT are important for the development of effective HIV/AIDS vaccine and passive immunization approaches to HIV-1. Indeed, MTCT of HIV is a unique situation where babies are exposed to the virus in presence of passively transferred antibodies.

This is why our studies focused on both the role of NAbs in MTCT of HIV-1 and the molecular characteristics of viruses transmitted from mothers to their infants. We restricted our analysis to HIV-1 CRF01_AE variants which belong to the predominant HIV-1 clade in Thai population, with the hope that our work would benefit to the development of strategies for prevention of HIV-1 infection, especially in Thailand.

2. Objectives of the entire study

The main purposes of the entire study were to define the role of maternal NAbs in MTCT of HIV-1 and to characterize the HIV-1 envelope glycoproteins of viruses preferentially transmitted in MTCT of HIV-1.

Specific aims were:

1. To determine the NAbs titers against heterologous HIV-1 primary isolates in transmitting and non-transmitting mothers, and to precise the molecular characteristics of CRF01_AE primary isolates that might be indicators of protective antibodies.
2. To examine the molecular characteristics of HIV-1 envelope glycoproteins of the viral variants from both mother and their children, considering the timing of transmission (*in utero* or intrapartum), trying to identify specific properties of transmitted viruses that might confer a selective advantage in the MTCT context.

3. Materials and methods, Results, Discussion

I. Maternal neutralizing antibodies against a CRF01_AE primary isolate are associated with a low rate of intrapartum HIV-1 transmission in Thailand

1. Background and specific aims

Maternal neutralizing antibodies (NAbs) could play a role in preventing HIV-1 transmission, particularly when the baby is exposed to the virus in the presence of optimal levels of Abs. Several studies have shown that non-transmitting mothers had more frequently detected and higher levels of NAbs responses than transmitting mothers [357; 227; 47; 156; 32], and the viruses transmitted intrapartum are escape variants resistant to maternal autologous NAbs [111; 410; 432]. In a previous study, we hypothesized that broadly cross-neutralizing, heterologous NAbs would protect babies against intrapartum HIV transmission [32]. We measured NAb titers against primary isolates of various clades in sera from pregnant Thai women for whom the time of transmission was known. We identified an association between higher titers of NAbs against a CRF01_AE primary isolate, MBA ,and lower rates of intrapartum transmission (Figure 31). However, only one isolate per clade was used in this previous study.

In the present study, we extended the analysis by using several CRF01_AE strains in a different Thai population selected on highly stringent criteria and confirmed the association we previously observed. In addition, we have identified a characteristic of the envelope glycoprotein of the MBA strain that might explain this association, at least in part.

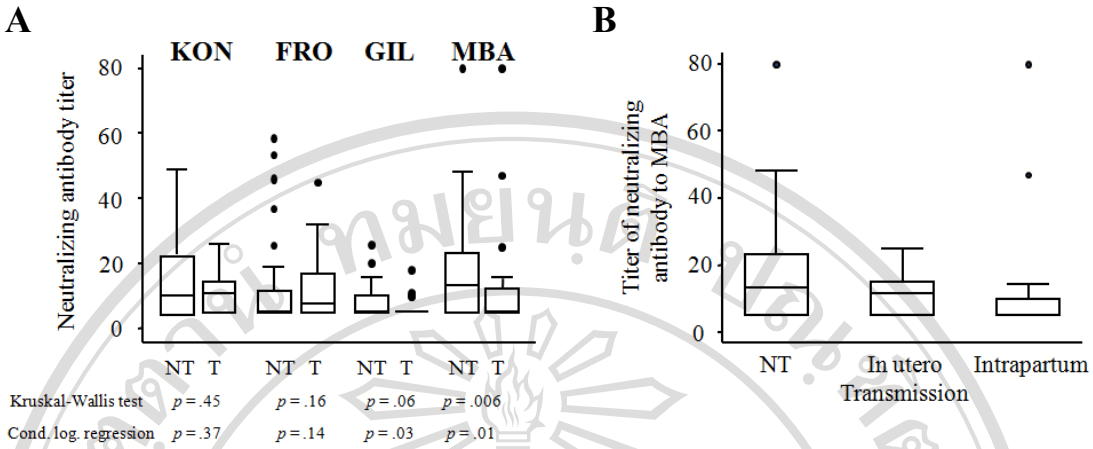


Figure 31. Comparison of neutralizing antibody titers in transmitting (T) versus non-transmitting (NT) mothers. (**A**) Antibodies to KON (CRF02_AG), FRO (B), GIL (F), and MBA (CRF01_AE) strains. (**B**) Antibodies to MBA strain in non-transmitting and transmitting mothers according to timing of transmission (*in utero* or intrapartum). The mothers who transmitted the virus intrapartum had significantly lower NAb titers to MBA, compared with non-transmitting mothers (Kruskal-Wallis test, $P = 0.008$; CLR, $P = 0.02$), whereas there was no significant difference between mothers who transmitted the virus *in utero* and those who did not (Kruskal-Wallis test, $P = 0.42$; CLR, $P = 0.53$). For each distribution, the horizontal lines represent the 25th, 50th (median), 75th, and 90th percentiles. (Figures from Barin *et al*. 2006 [32])

Specific aims were:

- to determine the NAb titers against heterologous primary isolates in transmitting and non-transmitting mothers, in order to confirm and extend previous findings.

- to characterize the molecular properties of HIV-1 strains that might be indicators of protective antibodies.

2. Materials and methods

Study population and sample collection

Samples were obtained from HIV-1-infected pregnant women enrolled in a clinical trial assessing various zidovudine (ZDV) treatment durations for the prevention of MTCT in Thailand (Perinatal HIV Prevention Trial, PHPT-1), in which infants were not breastfed [221]. The HIV-1 infection status of the infants was determined by HIV-1 proviral DNA PCR as previously described [32; 221]. We matched 45 transmitting mothers (cases) with samples available for this study to 45 non-transmitting mothers (controls) on baseline maternal viral load and duration of maternal ZDV prophylaxis, the two main independent baseline factors associated with

MTCT [185]. Table 2 displays the characteristics of these two groups. Of the 45 transmitting mothers, 14 mothers transmitted HIV-1 to their infants *in utero* and 29 mothers transmitted the virus intrapartum, and the timing of transmission was not determined in two cases. Blood samples were collected at baseline, before the start of ZDV prophylaxis, 4 to 11 weeks before delivery. Maternal viruses were subtyped, by both V3 serotyping [32] and phylogenetic analysis of the genomic sequences obtained after reverse transcriptase–polymerase chain reaction (RT-PCR) amplification of the gp41 region of the viral RNA in the plasma, as previously described [52]. Samples with discrepant results in V3 serotyping and genetic analysis were defined as indeterminate. HIV-1 CRF01_AE was identified in 80 (89%) of the 90 mothers and subtype B was identified in four (4%). Six were indeterminate.

Table 2. Baseline characteristics of transmitting and non-transmitting mothers

Characteristics	**Non-transmitters (n = 45)**	**Transmitters (n = 45)**	***P* value**
	Median (range)	Median (range)	Kruskal-Wallis test
Age (years)	26 (22-30)	25 (23-30)	0.61
CD4 cell count (per mm^3)	352 (240-434)	320 (220-450)	0.96
Viral load (log10 copies/mL)	4.38 (3.89-4.75)	4.41 (3.87-4.71)	0.97
Duration of zidovudine prophylaxis (weeks)	6 (4-11)	6 (4-11)	0.92
Hemoglobin level (g/dL)	10.6 (10.1-11.5)	10.4 (9.8-11.2)	0.23
Hematocrit (%)	32 (30-35)	32 (30-34)	0.33
Gestational age at sample acquisition (weeks)	26 (22-26)	26 (21-26)	0.60
Sample acquisition to delivery (weeks)	14 (13-17)	14 (12-19)	0.98

Heterologous primary isolates neutralization assay

Neutralization assays were carried out blind to the transmission status of the mother-infant pairs. The heterologous primary isolates selected for this study belonged to the two prevalent clades in Thailand (CRF01_AE and B) and presented various phenotypes (X4, R5, dual-tropic). There were three clade B strains, FRO (X4), BIG (R5) and CHA (R5X4), and three CRF01_AE strains, C1712 (X4), LEA (R5), MBA (R5X4). Virus stocks were prepared by a limited number of passages in PHA-stimulated peripheral blood mononuclear cells (PBMC). The same stock of each isolate was used for the entire study. The titers of NAbs against the six primary isolates in each maternal serum sample were determined using the P4P cell assay, as

previously described [32; 30]. P4P cells ($CD4^+CXCR4^+CCR5^+$ HeLa-derived cells harboring the *lacZ* reporter gene) were cultivated in DMEM supplemented with 10% FCS, 500 µg/ml G418, 100 µg/ml puromycin and 100 µg/ml antibiotics [74]. The assay was performed in duplicate and the results were expressed as mean values. A negative control was included in every series by use of a serum sample obtained from an HIV-seronegative, healthy individual without any risk factors. A pool of serum samples obtained from 10 untreated HIV-1-infected individuals was used for a positive control. The neutralization titer was defined as the reciprocal of the serum dilution resulting in a 90% decrease in the number of infected cells two days after infection with 100 $TCID_{50}$ (50% tissue culture infectious dose) when compared to the negative control.

Amplification and cloning of HIV-1 env genes

Genomic RNA from C1712, LEA and MBA viruses was extracted from each virus stock using the QIAamp® viral RNA Mini Kit (Qiagen, Hilden, Germany). Full-length (gp160) *env* gene was amplified by nested RT-PCR using subtype CRF01_AE *env*-specific primers. The outer primers pair was sensAEext (5'-GGTTARTTVARAGAATAAGAGAAAGAG-3') and asAEext (5'-TRCTTTTTGACCAYTTGCYYCCCAT-3'), and the inner primers pair was sensAEint (5'-AGAAGACAGTGGAAATGAGAGTGA-3') and asAEint (5'-ATDTTATRSCAAAGHCCTTTCDAAGCC-3'). Reverse transcription (RT) and the first round of PCR were carried out using the SuperScript™ One-Step RT-PCR for Long Templates kit (Invitrogen, Carlsbad, CA) under the following conditions: one RT of 30 minutes at 50°C, followed by the first round of PCR consisting in one cycle of 2 minutes at 94°C, 35 cycles of 15 seconds at 94°C, 30 seconds at 50°C, 3 minutes at 68°C, and an additional extension step of 10 minutes at 72°C. MBA *env* gene amplification differed from that of the two other viruses by the number of cycles (40 cycles) and the annealing temperature (45°C). A 5µl aliquot of the products of the first round of PCR was then used as template for the second round of amplification under the same cycling conditions, using Platinum® PCR SuperMix High Fidelity (Invitrogen). By this approach, PCR products of 2.6 kb were obtained for each virus. They were cloned into pCR2.1 (TOPO TA cloning® Kit; Invitrogen), before being excised by *Eco*RI restriction and transferred into the *Eco*RI site of the pCI expression vector (Invitrogen).

Sequence analysis

PCR products and pCR2.1-*env* clones were sequenced, using a set of *env*-specific internal primers, according to the Dye Terminator cycle sequencing protocol (Applied Biosystems, Foster City, CA). Nucleotide sequences were assembled, translated into amino acid sequences and aligned using the BioEdit package version 5.0.9 [158]. Potential N-linked glycosylation sites (PNGS) were identified by N-GlycoSite (http://www.hiv.lanl.gov/content.hiv-db/GLYCOSITE/glycosite.html)

[441]. MBA, C1712 and LEA *env* sequences have been submitted to GenBank and assigned accession numbers DQ518410 to DQ518412.

Construction of chimeric env genes

Two chimeric *env* genes, named V2LEA–*env*ΔV2MBA and V2MBA–*env*ΔV2LEA, were constructed in pCR2.1 vector by inverting V2 domains of LEA and MBA *env* genes using a domain exchange strategy as previously described [346]: the donor V2 domains of LEA and MBA were amplified by PCR using specific primers that anneal at both ends of V2 domains (Figure 32A). The primers pairs used for these amplifications were: LEAV2S (5'-TCTTTTAATATGACCACAGAACTAAAAGATAAG-3') and LEAV2AS (5'-ACAATTTATTAACCTATACTCACTATTATTACTACTTC-3') for LEA, and MBAV2S (5'-ACTTTTAATACGACCACAGAACTAGGAGAT-3') and MBV2AS (5'-ACAATGTATTAATATATACTTACTATAGGTACTATTGTTTTTCCC-3') for MBA. The recipient *env* backbones (deleted of the V2 domain) within the pCR2.1 vector were amplified by PCR using specific primers that anneal to regions adjacent to the V2 primers (Figure 32B). A 5' phosphate group (Phos) was added to these primers to allow ligation to the V2 amplicon. The primers pairs used for these amplifications were: LEAEnvS (5'-Phos-AATACTTCAGTCATTAAGCAGGCTTGTC-3') and LEAEnvAS (5'-Phos-ACAGTTTCTTACTTCATCTGTTATATTTCCTATTT-3') for LEA, and MBAEnvS (5'-Phos-AATTCTTCAGTCATTAAGCAAGCATGTCC-3') and MBAEnvAS (5'-Phos-GCAGTTTCTTACTGATTTTTCCATTTTTATTGTCTTAC-3') for MBA.

The PCR amplification conditions for the V2 domains were 1 cycle of 30 seconds at 98°C, followed by 35 cycles of 10 seconds at 98°C, 30 seconds at 57°C and 30 seconds at 72°C, and a final extension step of 72°C for 10 minutes. The amplification conditions for *env* backbones were the same except the extension time at 72°C that was increased to 4 minutes. PHUSION High Fidelity DNA Polymerase (New England Biolabs, Ipswich, MA) was used to generate blunt-ended PCR amplicons. PCR amplicons were digested with *Dpn*I to remove contaminating template DNA prior to ligation. Each donor V2 DNA fragment (V2 of LEA or MBA) was ligated to its recipient *env* backbone (pCR2.1–*env*ΔV2MBA or pCR2.1–*env*ΔV2LEA, respectively), to produce two pCR2.1 constructs containing chimeric *env* genes, V2LEA–*env*ΔV2MBA and V2MBA–*env*ΔV2LEA (Figure 32C). After transformation of *E. coli* DH5α by electroporation, individual bacterial colonies were screened for the presence and the correct orientation of the V2 domain into the *env* gene by enzymatic restriction analysis and sequencing of plasmid DNA. Correct chimeric *env* genes were then extracted from pCR2.1 by *Eco*RI restriction and transferred into the *Eco*RI site of the pCI expression vector (Invitrogen).

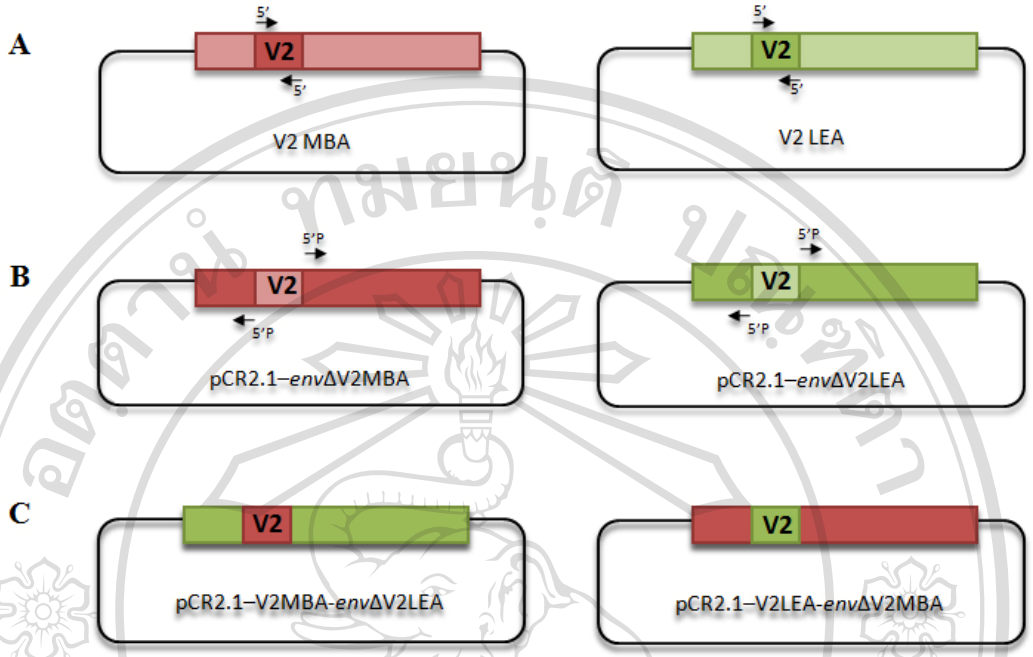


Figure 32. Construction of the V2 chimeric Envs. (**A**) The donor V2 domains (MBA and LEA) were amplified by PCR using specific primers that anneal at both ends of the V2 domain. (**B**) The recipient *env* backbones within the pCR2.1 vector (pCR2.1–*env*ΔV2MBA and pCR2.1–*env*ΔV2LEA) were amplified by PCR using specific primers that anneal to regions adjacent to the V2 primers. A 5' phosphate group was added to these primers to allow ligation to the V2 amplicon. (**C**) The V2 domains and recipient *env* backbones were blunt-end ligated to generate chimeric *env* genes in pCR2.1 (pCR2.1–V2MBA-*env*ΔV2LEA and pCR2.1–V2LEA-*env*ΔV2MBA).

Production and titration of envelope (Env)-pseudotyped viruses

293-T cells were cultivated in DMEM supplemented with 10% fetal calf serum (FCS) and 100 µg/ml antibiotics. TZM-bl cells ($CD4^+CXCR4^+CCR5^+$ HeLa-derived cells harboring the *lacZ* and *luc* reporter genes) were cultivated in DMEM supplemented with 10% FCS, 25 mM HEPES and 50 µg/ml gentamycin [315; 419]. Env-pseudotyped viruses were generated by co-transfecting 3 x 10^6 293T cells with 12 µg of each pCI-*env* plasmid and 8 µg of pNL4.3.LUC.R^-E^- [91], using calcium phosphate (Invitrogen). Briefly, at 18h after transfection, cells were washed and fresh medium was added. Two days later, viral supernatants were harvested, centrifuged, filtered (0,45 µm), aliquoted and frozen (-80°C). After determination of p24 antigen concentrations of viral stocks (INNOTEST® HIV Antigen mAb; Innogenetics, Belgium), quadruplicated serial five-fold dilutions of supernatants were used to infect TZM-bl cells in presence of 30 µg/mL DEAE-dextran. After 48 h, infection levels of target cells were determined by measuring luciferase activity present in cell lysates, using a quantitative luciferase assay (Bright-Glo luciferase assay; Promega) and a

Centro LB 960 luminometer (Berthold Technologies). The dilution of supernatant in which infection of target cells was observed in 50% of the wells ($TCID_{50}$) was determined for each virus. Infection was considered as positive in wells producing relative luminescence units (RLU) above 2.5-times background RLU.

Env-pseudotyped viruses neutralization assay

Pseudotyped viruses stocks were diluted to obtain 600 $TCID_{50}$/mL in growth medium. A volume of 25 µl corresponding to 15 $TCID_{50}$ was then incubated for 1h at 37°C with 75 µl of serial twofold dilutions of heat-inactivated serum in wells of 96-well culture plates. The virus/serum mixture was then used to infect 10,000 TZM-bl cells in presence of 30 µg/mL DEAE-dextran. After 48h, infection levels were determined by measuring luciferase activity as described above. The assay was performed in duplicate and the results were expressed as mean values. Neutralizing antibody titers were defined as the reciprocal of the serum dilution required to reduce RLU by 50%.

Statistical analysis

Categorical variables were compared using Fisher's exact test and a conditional logistic regression (CLR) model to take into account the matching of the two groups on maternal baseline viral load and ZDV prophylaxis treatment duration. Continuous variables were compared using the Wilcoxon matched-pairs signed-ranks test. We used McNemar's chi-squared test for matched case-control data to compare the proportions of patient sera able to neutralize each strain.

3. Results

Neutralizing activity of sera from mothers against six primary isolates

None of the 90 sera displayed no neutralizing activity, whereas 15 sera (17%) showed neutralizing activity against the six strains. Although cross-clade NAbs were detected in several sera, subtype-specific neutralizing antibodies predominated. Seventy six (95%) of the sera of the 80 CRF01_AE infected mothers neutralized at least one strain of subtype CRF01_AE, but only 50 (63%) neutralized at least one strain of subtype B. However, all CRF01_AE strains were not equally neutralized. The MBA strain was more resistant to neutralization than the other two CRF01_AE strains, LEA and C1712 (McNemar's test, $P < 0.001$ and $P = 0.001$, respectively): only 57 of 90 sera (63%) were able to neutralize MBA versus 81 (90%) for LEA and 75 (83%) for C1712. In contrast, the three subtype B isolates, BIG, CHA and FRO were similarly neutralized by 39 (43%), 35 (39%) and 38 (42%) of 90 sera, respectively (McNemar's test not significant).

The percentage of mothers with detectable NAbs against each of the three subtype B isolates and two of the CRF01_AE isolates (LEA and C1712) was not significantly different in transmitting and non-transmitting mothers (Table 3). In contrast, NAbs against MBA, the CRF01_AE strain the most resistant to neutralization, were detected in 34 of 45 non-transmitting mothers (76%), versus only 23 of 45 (51%) in transmitting mothers (Fisher's exact test: $P = 0.03$; CLR: $P = 0.02$, odds ratio 3.8; 95% confidence interval: 1.2 to 11.3) (Table 3). This difference remained significant if the analysis was restricted to the 80 CRF01_AE infected mothers (CLR: $P = 0.01$, odds ratio 6.5; 95% confidence interval: 1.5 to 28.8). Similarly, the distribution of titers against BIG, CHA, FRO, LEA and C1712 did not differ significantly between transmitting and non-transmitting mothers, whereas higher titers against MBA were observed in non-transmitting mothers (Wilcoxon signed-rank, $P = 0.01$) (Figure 33A). This association remained significant when the analysis was restricted to the CRF01_AE infected mothers ($P = 0.009$).

Neutralizing antibodies against MBA were undetectable in 15 of the 29 mothers who transmitted the virus intrapartum (52%) versus only 7 of the 29 (24%) matched controls (Fisher's exact test, $P = 0.06$; CLR; $P = 0.05$), and were undetectable in six (43%) of the 14 mothers who transmitted the virus *in utero* versus two of the 14 (14%) matched controls (Fisher's exact test, $P = 0.21$; CLR, $P = 1.00$). Similarly, the levels of neutralizing antibodies against MBA were significantly lower in mothers who transmitted the virus intrapartum, compared to their matched controls (Wilcoxon signed-rank, $P = 0.03$). In contrast they were not statistically different between mothers who transmitted the virus *in utero* and their matched controls (Wilcoxon signed-rank, $P = 0.15$) (Figure 33B).

Table 3. Comparison of detectable neutralizing antibodies against the 6 primary isolates in transmitting and non-transmitting mothers

Presence of detectable NAbs against:	**Nontransmitters (n = 45)**	**Transmitters (n = 45)**	***P* value**	
	n (%)	n (%)	Fisher's exact test	CLR
BIG (B/R5)	22 (49)	17 (38)	0.40	0.30
CHA (B/R5X4)	18 (40)	17 (38)	1.00	0.84
FRO (B/X4)	20 (44)	18 (40)	0.83	0.67
LEA (CRF01_AE/R5)	41 (91)	40 (89)	1.00	0.74
MBA (CRF01_AE/R5X4)	34 (76)	23 (51)	**0.03**	**0.02**
C1712 (CRF01_AE/X4)	37 (82)	38 (84)	1.00	0.78

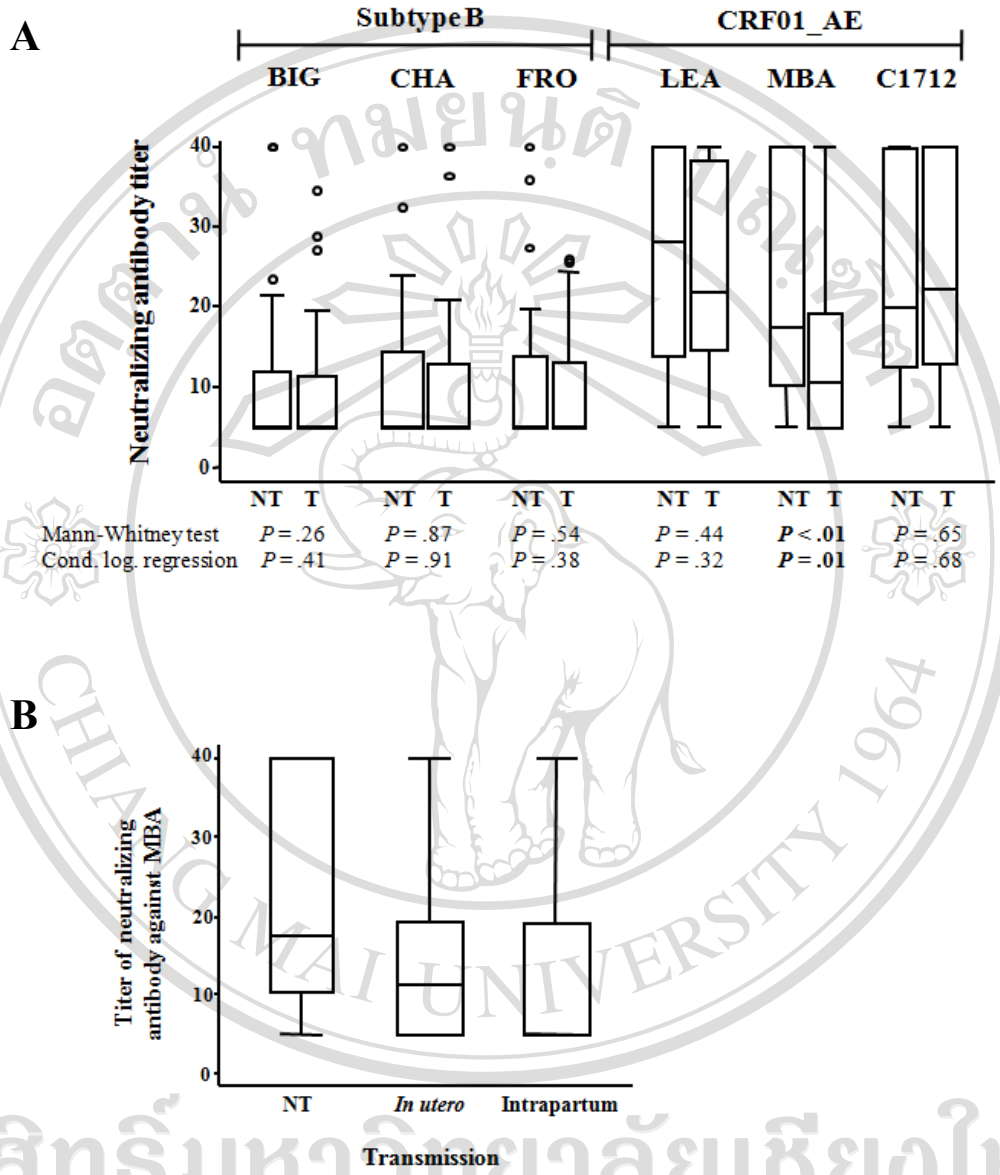


Figure 33. Comparison of neutralizing antibody titers in transmitting (T) and non-transmitting (NT) mothers. (**A**) Comparison of NAb titers against the six strains in transmitting and non-transmitting mothers. Box plots show the distribution of maternal antibody titers; for each distribution, the horizontal lines represent the lower adjacent 25th, median, 75th, and upper adjacent percentiles. **(B)** Comparison of NAb titers against the MBA strain in non-transmitting mothers and transmitting mothers according to time of transmission (*in utero* or intrapartum).

Molecular characteristics of the envelope glycoprotein of MBA

We tried to identify the molecular characteristics of the envelope glycoprotein of MBA that might provide at least partial explanations for the association between lower risk of MTCT and NAbs against this isolate. We focused on the three CRF01_AE strains. Alignments of the amino-acid Env sequences from the three strains revealed one major difference between MBA and the other two strains (Figure 34). The MBA Env showed an unusually long V2 domain of 63 amino acids (23 and 25 amino acids longer than that of LEA and C1712, respectively) encoding 6 PNGS (*versus* 2 PNGS in LEA and C1712). Using the HIV Blast program, searches of homology between the MBA Env and sequences found in the HIV sequence database (http://www.hiv.lanl.gov/content/index) showed that the V2 lengthening of MBA is exceptional. A detailed analysis of the V2 domains of the 68 CRF01_AE strains present in the web-alignment of the full spectrum of HIV sequences (http://www.hiv.lanl.gov/content/sequence/NEWALIGN/align.html) contained in the HIV sequence database in 2006, revealed that the longer V2 domain was composed of 53 amino acids (accession number U51188) and the more potentially glycosylated V2 domain encoded 5 PNGS (accession number AY444805). In addition to this major particularity, 3 PNGS localized in the C2 constant domain, the V4 variable loop and the membrane proximal external region (MPER) of gp41, respectively, as well as a deletion of 7 amino-acids in the gp41 cytoplasmic tail, were observed in MBA but not in LEA and C1712. Nevertheless, the three PNGS observed in MBA were not specific of this strain. They were found in the envelope of the highly sensitive to neutralization strain HxB2 (for the two PNGS in C2 and MPER) or in envelopes of several other CRF01_AE strains (for the two PGNS in V4 and MPER) (data not shown).

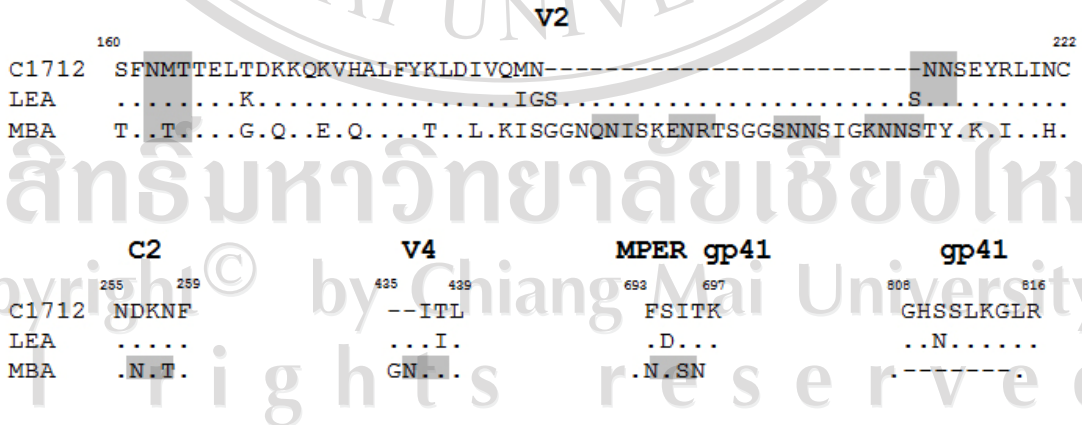


Figure 34. Locations of the sequence characteristics of the envelope glycoprotein of MBA. The amino acid Env sequences of C1712, LEA and MBA were aligned using the BioEdit package version 5.0.9 [158]. Only regions showing characteristics specific to MBA are indicated. Amino acid numbering is based on MBA amino acid Env sequence. Identical amino acids and insertions are indicated by dots or dashes, respectively. PNGS (NXT or NXS) are highlighted in grey.

This analysis led us to hypothesize that the unusual long V2 domain might contribute to render MBA more resistant to neutralization than LEA and C1712. To test this hypothesis, pseudotyped viruses expressing wild type Envs of MBA and LEA and two chimeric Envs constructed by swapping the V2 domains (V2LEA–EnvΔV2MBA and V2MBA–EnvΔV2LEA) were produced. The chimeric V2LEA-EnvΔV2MBA pseudotyped virus, expressing the V2 domain of LEA in an MBA Env backbone, was not infectious. The neutralization profiles of the three other viruses, i.e. MBA, LEA and V2MBA–EnvΔV2LEA were compared, using 10 serum samples of mothers selected among those that neutralized more efficiently LEA than MBA (Figure 35). As shown on figure 35 and table 4, the V2 domain of MBA conferred to LEA a higher level of resistance to neutralization by 6 of the 10 sera. Indeed, the chimeric virus displayed profiles of neutralization by sera WT1064 and WT1382 similar to those of MBA, with a decrease in neutralization sensitivity of approximately threefold when compared to LEA. The neutralization profiles of the chimeric virus by sera WT0744, WT0769, WT0836 and WT1388 were intermediate between those of LEA and MBA, with decreases in sensitivity ranging from 1.5- to 2.2-fold when compared to LEA. In contrast, the V2 region of MBA did not influence the neutralization profile of the chimeric virus by sera WT0069, WT0665, WT0821 and WT1201. These results suggest that the V2 region of MBA certainly contributes to the neutralization resistance of this strain, but that other determinants might be involved.

Table 4. Neutralizing antibody titers of 10 serum samples against LEA, MBA and V2MBA-EnvΔV2LEA pseudotyped viruses. Results are expressed as the reciprocal dilution that reduces the RLU by 50%.

Serum	LEA	MBA	V2MBA–EnvΔV2LEA
WT1064	100	32.5	44.2
WT1382	95	30.5	28.1
WT0744	98.7	46.8	66.8
WT0769	193.2	54.6	87
WT0836	146.1	38.1	89.8
WT1388	34	12.7	21.6
WT0069	981	572.9	1014.7
WT0665	182.9	22.3	183.2
WT0821	108.7	57.3	93.5
WT1201	217.1	65.1	240.3

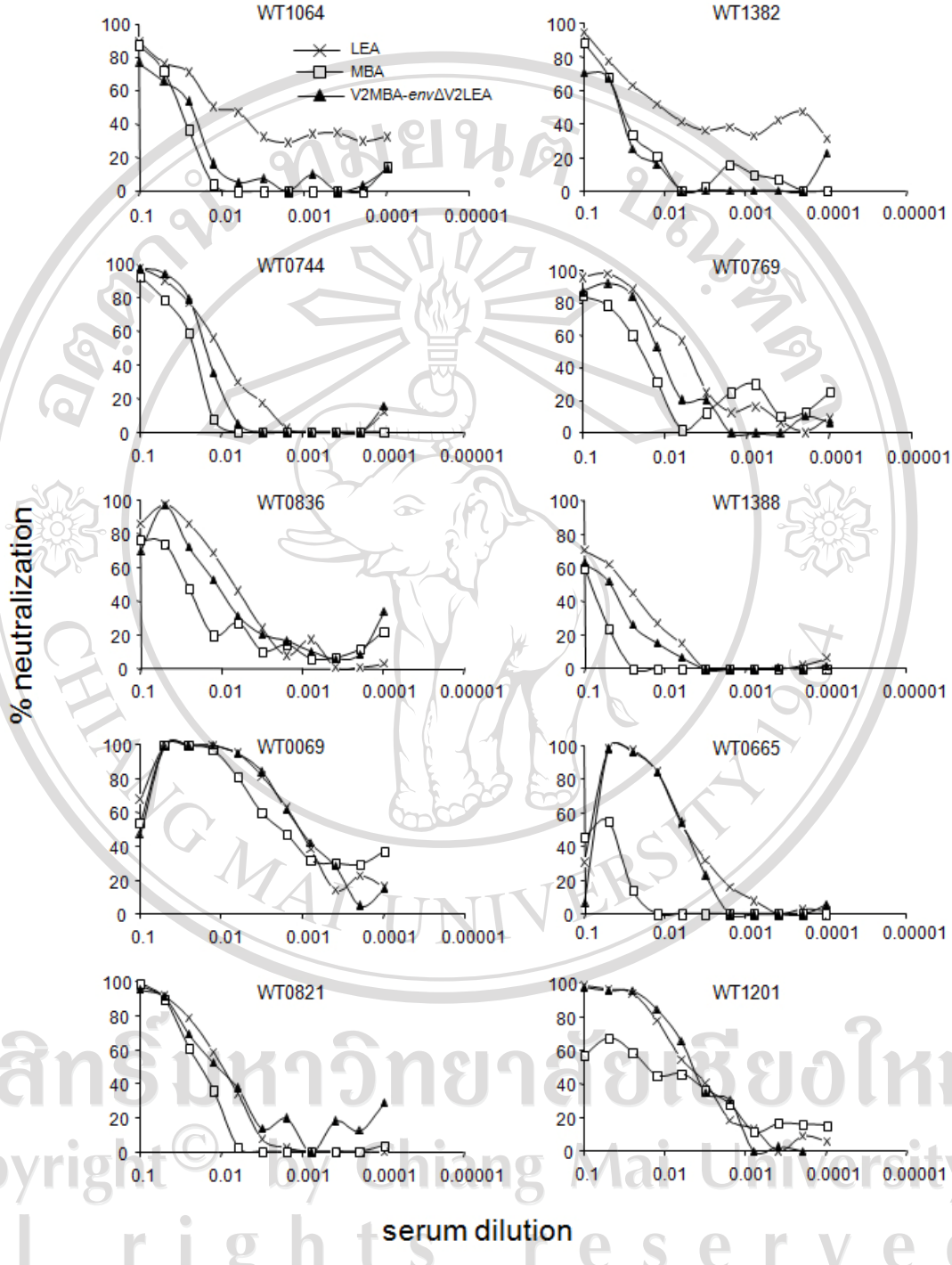


Figure 35. Effect of the V2 domain of MBA on the neutralization of LEA. Neutralization activities of 10 mothers sera were determined against viruses pseudotyped with the parental Envs of LEA (crosses) and MBA (open squares), and with the chimeric envelope V2MBAEnvΔV2LEA (closed triangles). Percentages of neutralization are plotted against the serum dilutions. The sera used are indicated at the top of each panel.

4. Discussion

We found an association between high titers of maternal NAbs against the MBA strain — a primary isolate of the prevalent clade in Thailand — and a lower rate of mother-to-child transmission of HIV, specifically for intrapartum transmission. These findings confirm the results of a previous study of a random sample of 28 transmitting and 62 non-transmitting mothers [32]. Therefore, we conclude that such antibodies may be associated with a lower risk of MTCT. As intrapartum exposure to HIV resembles natural challenge in the presence of preexisting antibodies, it might be possible to identify correlates of protection useful for ongoing or future vaccine studies. In this study, neutralizing activity was tested against six primary isolates of different phenotypes. Three of these strains belong to the predominant clade in Thailand, CRF01_AE. The other three belong to the less prevalent clade B. Among CRF01_AE strains, MBA showed the particularity to be more resistant to neutralization than the two other strains. The three clade B strains were similarly neutralized. No association between neutralizing antibodies and MTCT was found for the three B strains and for two of the CRF01_AE strains (LEA and C1712). In contrast, the proportion of mothers who harbored NAbs against the MBA strain was significantly higher for non-transmitting than for transmitting mothers, and NAb titers against the MBA strain were significantly higher in non-transmitters than in transmitters. Higher levels of NAbs against the MBA strain were associated significantly with lower rates of intrapartum transmission, but not of *in utero* transmission. This suggests that only NAbs able to neutralize strains with specific properties, such as MBA, might be indicators for protection.

We therefore investigated the molecular characteristics of the envelope gene of the MBA strain. We found that MBA showed an exceptionally long V2 domain of 63 amino acids including 6 PNGS. Several reports have shown that the V1V2 domain strongly influences sensitivity to neutralization by masking neutralizing targets [346; 313; 350; 370; 429]. We therefore hypothesized that this unusual V2 domain might contribute to render MBA more resistant to neutralization when compared to the other two CRF01_AE isolates, LEA and C1712. To test this hypothesis, we compared the neutralization profiles of pseudotyped viruses expressing either wild type Envs of MBA or LEA or a chimeric Env containing the V2 domain of MBA in a LEA Env backbone. We found that the V2 region of MBA increased the resistance of LEA to neutralization by 6 of the 10 tested sera, indicating that the unusual V2 domain contributed certainly to the neutralization resistance of MBA, but was probably not the only determinant involved. This observation would suggest that the antibodies effective in reducing the risk of MTCT would be potent antibodies targeting either epitopes exposed on the large V2 domain itself or, alternatively, distant epitopes specifically exposed because of a particular conformation imposed by this large heavily glycosylated structure. Interestingly, Pinter *et al* recently showed that the V1V2 domain is a global regulator of sensitivity of primary isolates to neutralization by antibodies commonly induced upon infection but not by exceptional antibodies capable of neutralizing a broad range of primary isolates, like the human monoclonal antibodies 2F5, 2G12 or b12 [313]. Further characterization of maternal antibodies

with potent neutralizing activity against MBA and associated with a lower risk of MTCT, may therefore allow the identification of key neutralization epitopes. Such epitopes may constitute additional potential targets for the development of a protective vaccine against HIV.

In conclusion, this study confirms that higher titers of maternal antibodies against a CRF01_AE primary isolate, MBA, are associated with a lower intrapartum risk of HIV-1 transmission in Thailand, and confirms a major implication of the V2 domain of gp120 in the neutralization process. The results suggest that certain primary isolates might be indicative of a protective antibody response. Further studies aiming to select strains indicative of neutralization/protection might facilitate the identification of correlates of protection crucial for the development of protective vaccines. Lessons from the MTCT model would be very useful in this respect.

5. Publications

These data have been submitted to the Journal of Infectious Diseases in August, 2008.

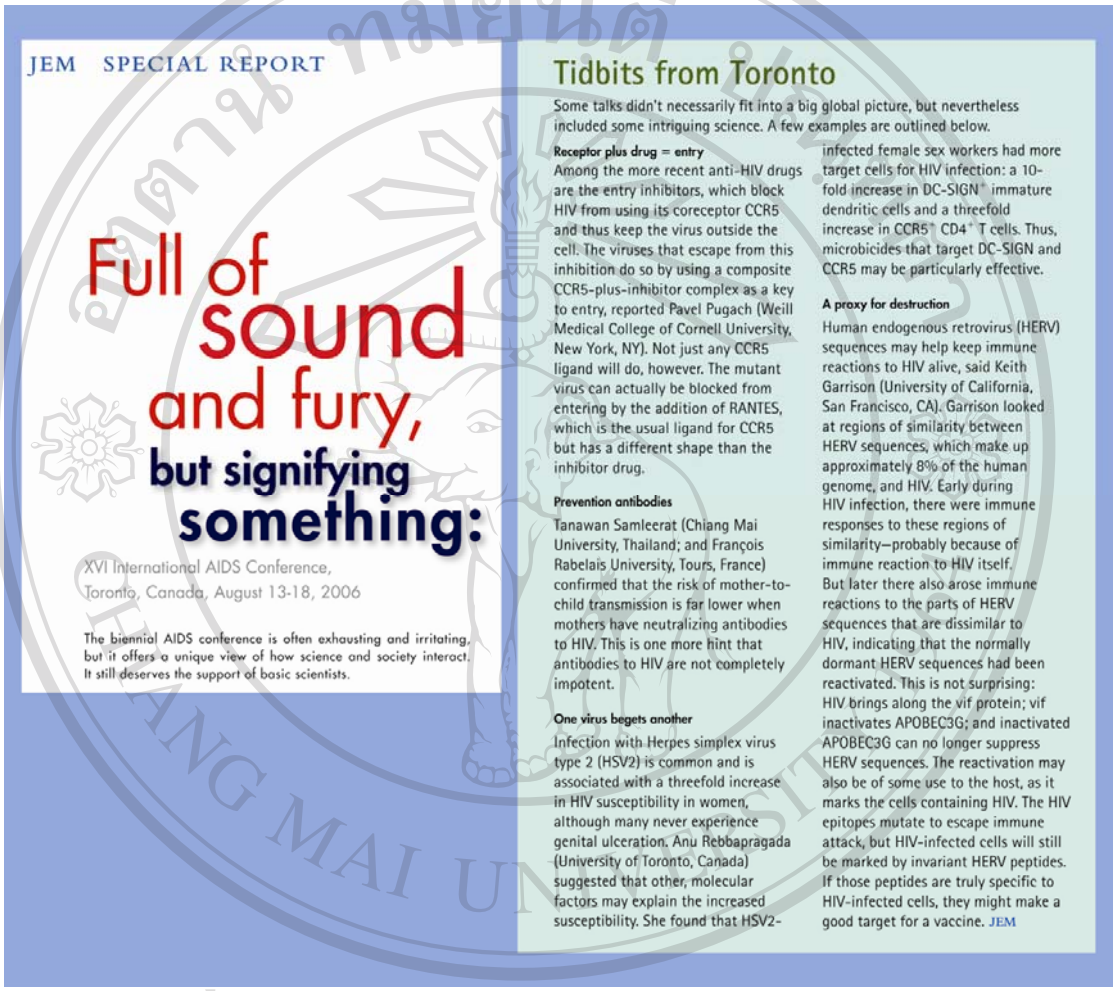
> **Samleerat T**, Thenin S, Jourdain G, Ngo-Giang-Huong N, Moreau A, Leechanachai P, Sirithadthamrong S, Surasaerneewongse V, Warachit B, Hotrawarikarn S, Lallemant M, Barin F, and Braibant M. *Maternal neutralizing antibodies against a CRF01_AE primary isolate are associated with a low rate of intrapartum HIV-1 transmission in Thailand*.

This work was presented in part at the XVI International AIDS Conference, Toronto, Canada, August 13-18, 2006 and at the AIDS Vaccine 2006 Conference, Amsterdam, The Netherlands, August 29-September 1, 2006.

> **Samleerat T**, Jourdain G, Braibant M, Ngo-Giang-Huong N, Lallemant M, Leechanachai P, Sirithadthamrong S, Surasaerneewongse V, Warachit B, Hotrawarikarn S, and Barin F. *Maternal neutralizing antibodies to a CRF01_AE primary isolate are associated with low intra-partum transmission of HIV-1 in Thailand*. The XVI International AIDS Conference, Toronto, Canada, August 13-18, 2006 (***oral presentation-presenting author***).

> **Samleerat T**, Jourdain G, Braibant M, Ngo-Giang-Huong N, Lallemant M, Leechanachai P, Sirithadthamrong S, Surasaerneewongse V, Warachit B, Hotrawarikarn S, and Barin F. *Maternal neutralizing antibodies toward a CRF01_AE primary isolate are associated with lower mother-to-child transmission of HIV-1 in Thai women*. The AIDS Vaccine 2006 Conference, Amsterdam, The Netherlands, August 29-September 1, 2006 (***poster presentation-presenting author***).
>
> Abstract published in: *Antiviral Therapy, 2006; Suppl 2: 1-251.*

The work presented at the XVI International AIDS Conference, Toronto, Canada was remarked in the special report of Journal of Experimental Medicine, 2006; 203 (11): 2394-2403.



JEM SPECIAL REPORT

Full of sound and fury, but signifying something:

XVI International AIDS Conference, Toronto, Canada, August 13-18, 2006

The biennial AIDS conference is often exhausting and irritating, but it offers a unique view of how science and society interact. It still deserves the support of basic scientists.

Tidbits from Toronto

Some talks didn't necessarily fit into a big global picture, but nevertheless included some intriguing science. A few examples are outlined below.

Receptor plus drug = entry

Among the more recent anti-HIV drugs are the entry inhibitors, which block HIV from using its coreceptor CCR5 and thus keep the virus outside the cell. The viruses that escape from this inhibition do so by using a composite CCR5-plus-inhibitor complex as a key to entry, reported Pavel Pugach (Weill Medical College of Cornell University, New York, NY). Not just any CCR5 ligand will do, however. The mutant virus can actually be blocked from entering by the addition of RANTES, which is the usual ligand for CCR5 but has a different shape than the inhibitor drug.

Prevention antibodies

Tanawan Samleerat (Chiang Mai University, Thailand; and François Rabelais University, Tours, France) confirmed that the risk of mother-to-child transmission is far lower when mothers have neutralizing antibodies to HIV. This is one more hint that antibodies to HIV are not completely impotent.

One virus begets another

Infection with Herpes simplex virus type 2 (HSV2) is common and is associated with a threefold increase in HIV susceptibility in women, although many never experience genital ulceration. Anu Rebbapragada (University of Toronto, Canada) suggested that other, molecular factors may explain the increased susceptibility. She found that HSV2-infected female sex workers had more target cells for HIV infection: a 10-fold increase in DC-SIGN$^+$ immature dendritic cells and a threefold increase in CCR5$^+$ CD4$^+$ T cells. Thus, microbicides that target DC-SIGN and CCR5 may be particularly effective.

A proxy for destruction

Human endogenous retrovirus (HERV) sequences may help keep immune reactions to HIV alive, said Keith Garrison (University of California, San Francisco, CA). Garrison looked at regions of similarity between HERV sequences, which make up approximately 8% of the human genome, and HIV. Early during HIV infection, there were immune responses to these regions of similarity—probably because of immune reaction to HIV itself. But later there also arose immune reactions to the parts of HERV sequences that are dissimilar to HIV, indicating that the normally dormant HERV sequences had been reactivated. This is not surprising: HIV brings along the vif protein; vif inactivates APOBEC3G; and inactivated APOBEC3G can no longer suppress HERV sequences. The reactivation may also be of some use to the host, as it marks the cells containing HIV. The HIV epitopes mutate to escape immune attack, but HIV-infected cells will still be marked by invariant HERV peptides. If those peptides are truly specific to HIV-infected cells, they might make a good target for a vaccine. JEM

II. Characteristics of human immunodeficiency virus type 1 (HIV-1) gp120 *env* sequences in mother-child pairs infected with HIV-1 subtype CRF01_AE

1. Background and specific aims

HIV-1 can rapidly escape from autologous NAbs responses [16; 343; 420]. Evolution of both potential N-linked glycosylation sites (PNGS), which anchor a "glycan shield" of carbohydrates to the envelope, and variable loops, may contribute to escape from NAbs. Consistent with this hypothesis, recent studies of heterosexually transmitted viruses identified soon after infection have suggested that these early viruses have fewer PNGS and shorter variable loop sequences in the envelope than later isolates, depending on subtype [107; 79; 132; 236], indicating that smaller, less highly glycosylated Envs may be favored for transmission. As with sexual transmission, a bottleneck occurs during MTCT. As reviewed above, previous studies have demonstrated that HIV-1-infected mothers generally have a heterogeneous viral population, whereas their infected infants have a more homogeneous virus population [428; 357; 13; 301; 410]. It was suggested that *in utero* transmissions were associated more likely with single or multiple maternal variants and intrapartum transmissions more likely with minor HIV-1 variants [111]. However, this was not confirmed in other studies [410; 109]. The most extensive molecular studies published so far included a maximum of 12 to 14 mother-infant pairs, were conducted in populations infected with different subtypes, and used different molecular approaches (heteroduplex mobility assay and/or phylogenetic analysis after sequencing [410; 109; 432]). In most studies in which the *env* gene was sequenced, only portions of the gp120 region were analyzed, with only one study comparing the entire V1-to-V5 region of 78 clones from 12 pairs [432]. In contrast with these previous studies, we investigated the molecular characteristics of the HIV-1 variants involved in mother-to-child transmission by focusing on a homogeneous population infected with CRF01_AE isolates in Thailand, analyzing entire gp120 *env* sequences (353 clones) from 17 mother-child pairs for which the timing of transmission was known.

Specific aim was to examine the molecular characteristics of HIV-1 envelope glycoprotein of the virus isolates from mother and their children, considering timing of transmission (*in utero* or intrapartum), looking for molecular characteristics that might confer a selective advantage for MTCT.

2. Materials and methods

Study population and sample collection

Samples were obtained from participants enrolled in a clinical trial for the prevention of MTCT in Thailand (Perinatal HIV Prevention Trial, PHPT-1) [221]. The infants were not breastfed. Their HIV-1 infection status was determined by PCR-based HIV-1 proviral DNA detection, using DNA extracted from peripheral blood collected within 48 h of birth, at 6 weeks, and at 4 and 6 months of age. Infants were considered to be uninfected if two PCR tests for HIV DNA were negative on two separate occasions after one month of age. Infants were considered to be infected with HIV if the test was positive on two separate occasions. We considered the infant to have been infected *in utero* if the first test result, using blood collected within 48 h of birth, was positive, and intrapartum if the first test result was negative but the following test results were positive [56]. We randomly selected 24 HIV-1-infected mother-child pairs for whom samples were available. We analyzed maternal peripheral blood taken at the time of delivery and plasma from the infant taken at, or as close as possible to the first time point at which HIV-1 DNA PCR was positive. The median time of sampling for infant plasma samples was 44 days after delivery (interquartile range, 39 to 49 days). The median plasma viral load (VL) in infants was 4.7 $\log_{10}$ (range, 2.2 – 5.1). We focused on viral sequences corresponding to the proviral DNA present in the peripheral blood mononuclear cells (PBMCs) of the mothers, rather than the RNA present in their plasma virions, to obtain more complete information about the evolution of the viral population during infection. Furthermore, the chemoprophylaxis administered to the mothers frequently decreased viral load to levels too low for amplification of the entire gp120 *env* fragment from viral RNA. In contrast, we focused on RNA corresponding to plasma virions from the infants, for description of the genetic characteristics of the early viral population that was successfully transmitted. All experiments were performed blind to the timing of transmission.

Amplification and cloning of HIV-1 env genes

Genomic DNA was extracted from maternal peripheral blood, using the QIAamp DNA Blood Midi Kit (Qiagen). Viral RNA was extracted from infant plasma samples (280 µl), using the QIAamp Viral RNA Mini Kit (Qiagen), and cDNA was synthesized from the entire extracted RNA with the SuperScript First-Strand Synthesis System for RT-PCR (Invitrogen), using the specific primer AEa1ext (5'-TTCTCTTTGCCYTGGTGGGTGCTA-3'). A 1.2 kb fragment of the V1-to-V5 region covering almost the entire HIV-1 *env* gp120 gene (from upstream V1 to downstream V5) was amplified by nested PCR (Platinum PCR SuperMix High Fidelity; Invitrogen), using subtype-specific primers. The primers were based on an alignment of sequences of CRF01_AE isolates obtained from the HIV databases at Los Alamos National Laboratory (https://hiv.lanl.gov). The primers AEs1ext (5'-TGGGTTACAGTTTATTATGGGG-3') and AEa1ext were used in the first round of

amplification, and the primers AEs1int (5'-TGCCAAAGCATATGAGACAGARGYGCA-3') and AEa1int (5'-TTCACTTCTCCAATTGTCCTTYATRTT-3') were used in the second round. Plasma virus-derived cDNA (2 µl corresponding to 10% of the 1st strand reaction) or proviral DNA (1µg) was used as a template for the first-round PCR and were amplified under the following conditions: heating at 94ºC for 2 min, 35 cycles of 94ºC for 30 s, 50ºC for 30 s, and 68ºC for 90 s, and a final extension step of 72ºC for 10 min. The first PCR product was purified with the NucleoSpin Extract II Kit (Macherey-Nagel, Germany), to adjust the total volume of PCR product to 10 µl, for use as a template for the second round of PCR, under the same cycling conditions. PCR amplification were performed separately for each individual sample, to avoid cross-subject contaminations. PCR products were obtained from both mother and infant of 17 pairs. We therefore focused on these 17 pairs.

The 1.2 kb PCR products were gel purified with the NucleoSpin Extract II Kit and inserted into the pCR2.1-TOPO TA cloning vector (Invitrogen, Carlsbad, CA). Individual bacterial colonies were screened for the presence of inserted *env* gene fragments by restriction enzyme analysis of the plasmid DNA.

Sequencing and genetic analysis

Eight to 15 positive clones were sequenced for each subject. HIV-1 *env* clones were sequenced with the Big Dye Terminator V3.1 Cycle Sequencing kit (Applied Biosystems), using six overlapping primers encompassing the entire gp120 sequence. The sequencing products were analyzed on an ABI 3100 genetic analyzer (Applied Biosystems).

Nucleotide sequences were assembled and translated into amino-acid sequences with the BioEdit package [158]. Nucleotide and deduced amino-acid sequences were aligned with Clustal W [394] and manually adjusted to optimize the alignment. Sequences with premature stop codons were excluded. The amino-acid numbering system used here was that of the HIV-1 CFR01_AE consensus sequence obtained from the HIV database (http://hiv.lanl.gov). The HIV-1 subtype references were obtained from the HIV databases at Los Alamos National Laboratory and were used to determine the HIV-1 subtype of all nucleotide sequences from the 17 mother-child pairs. Neighbor-joining trees were constructed, using nucleotide sequences spanning the V1-to-V5 region and MEGA version 3.1 [214]. Branching-order reliability was assessed by bootstrap analysis with 1,000 replicates. Potential N-linked glycosylation sites (PNGS) were identified with N-Glycosite (https://hiv.lanl.gov/content/hiv-db/GLYCOSITE/glycosite.html). Genetic distances were calculated with the Kimura two-parameter method (transition-to-transversion ratio of 2.0) [198].

Statistical analysis

We used the Wilcoxon signed-rank test to compare the intra-patient genetic distances between the mother and infant sequences, overall and according to the

timing of transmission, and the Kruskall-Wallis test between *in utero* and intrapartum cases of transmission among mothers or infants. To study the differences between length and number of PNGS of the mothers' and infants' sequences, taking into account their distribution in the mother and in the infant, we used a linear mixed-effects model fit by maximum restricted likelihood, where mother-infant pairs were considered as a random effect and the variable "mother or infant" as both a fixed and a random effect.

In an exploratory analysis, we fitted an exact logistic regression model conditional on the number of PNGS within pairs (Procedure Exlogistic, Stata 10) to estimate for each position the ratio of the odds of the presence of a PNGS in infants' clones to the odds of it in mothers' clones, and rank the amino acid positions in the gp120 accordingly. To take into account the multiple comparisons, we considered only differences significant at $P \leq 0.001$.

Nucleotide sequence accession numbers

All sequences (353 V1-to-V5 *env* fragments) were deposited in GenBank under accession numbers EU031053 to EU031405.

3. Results

Characteristics of the patients

Among the 17 pairs, six infants were considered to have been infected *in utero* and 11 were considered to have been infected intrapartum.

Phylogenetic analysis of HIV-1 env sequences from mother-child pairs

We analyzed 353 envelope clones covering almost the entire gp120 region of the *env* gene (175 clones from mothers and 178 clones from infants). All 353 V1-to-V5 sequences clustered with CRF01_AE reference sequences, with a high bootstrap value (data not shown), indicating that all cases were infected with viruses of the HIV-1 CRF01_AE clade. A neighbor-joining tree constructed for the 353 V1-to-V5 *env* nucleotide sequences from the 17 mother-child pairs showed that the sequences from each pair formed a monophyletic group separated from those of the other mother-child pairs, indicating an epidemiological link between each mother and her child and an absence of cross-sample contamination (Figure 36). In addition, the inter-pair genetic distance (median, 14.5%; IQR 12.8 to 16.3%) was significantly higher ($P<$ 0.01) than the intra-pair genetic distance (median, 3.1%; IQR 2.1 to 5.8%; Figure 37), indicating that *env* nucleotide sequences of the epidemiologically linked mother-child pairs were more similar to each other than to those of epidemiologically unlinked individuals.

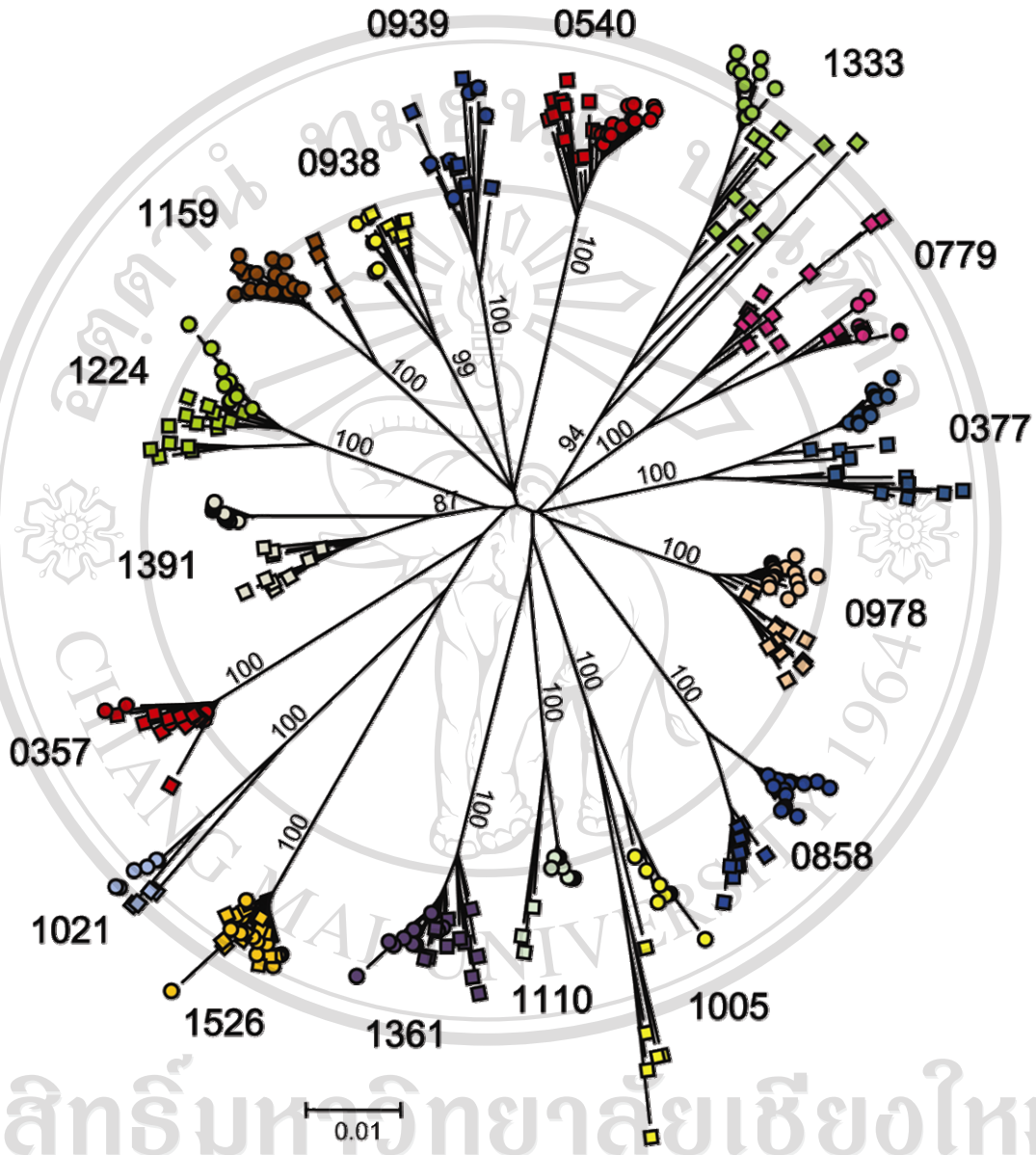
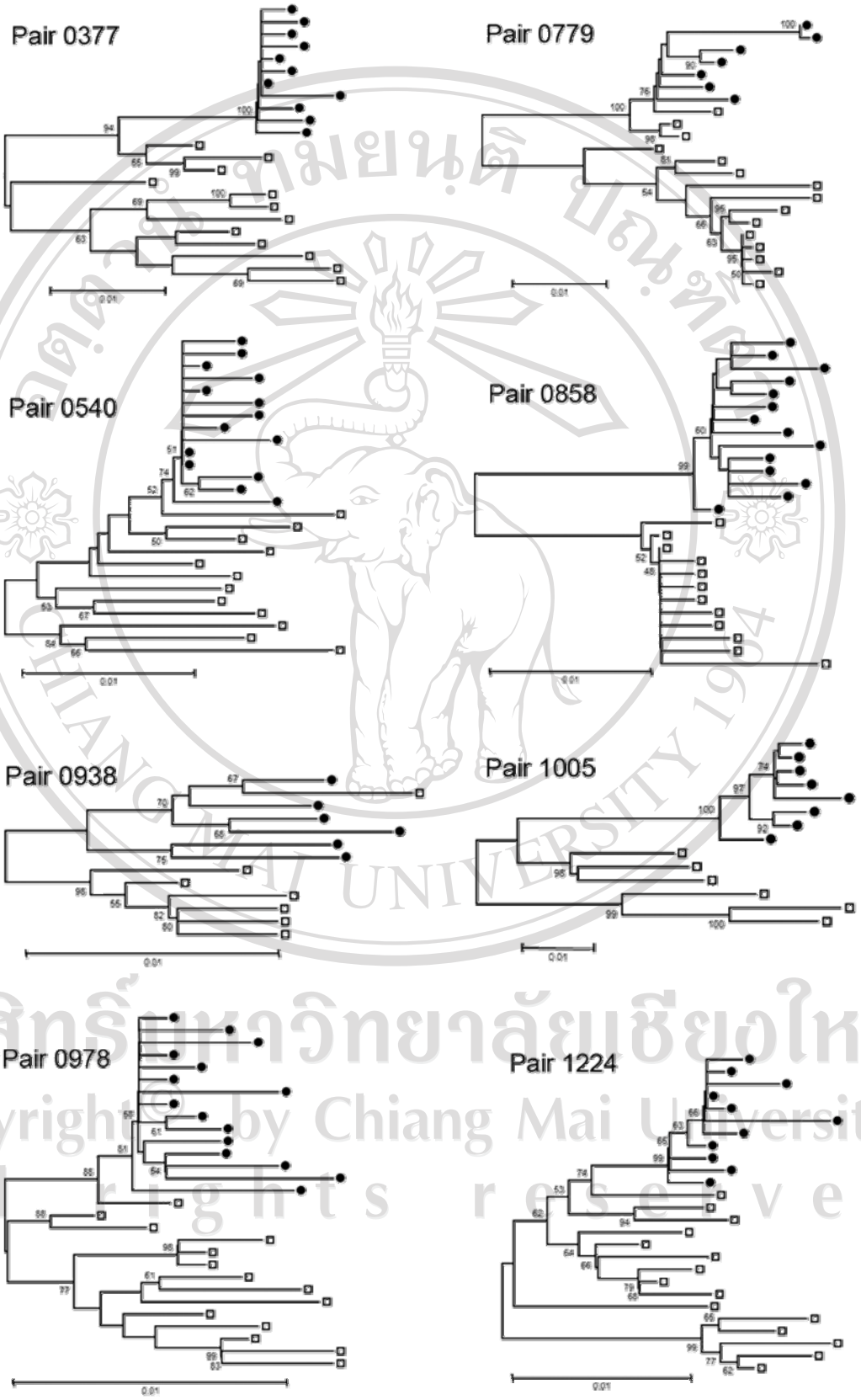
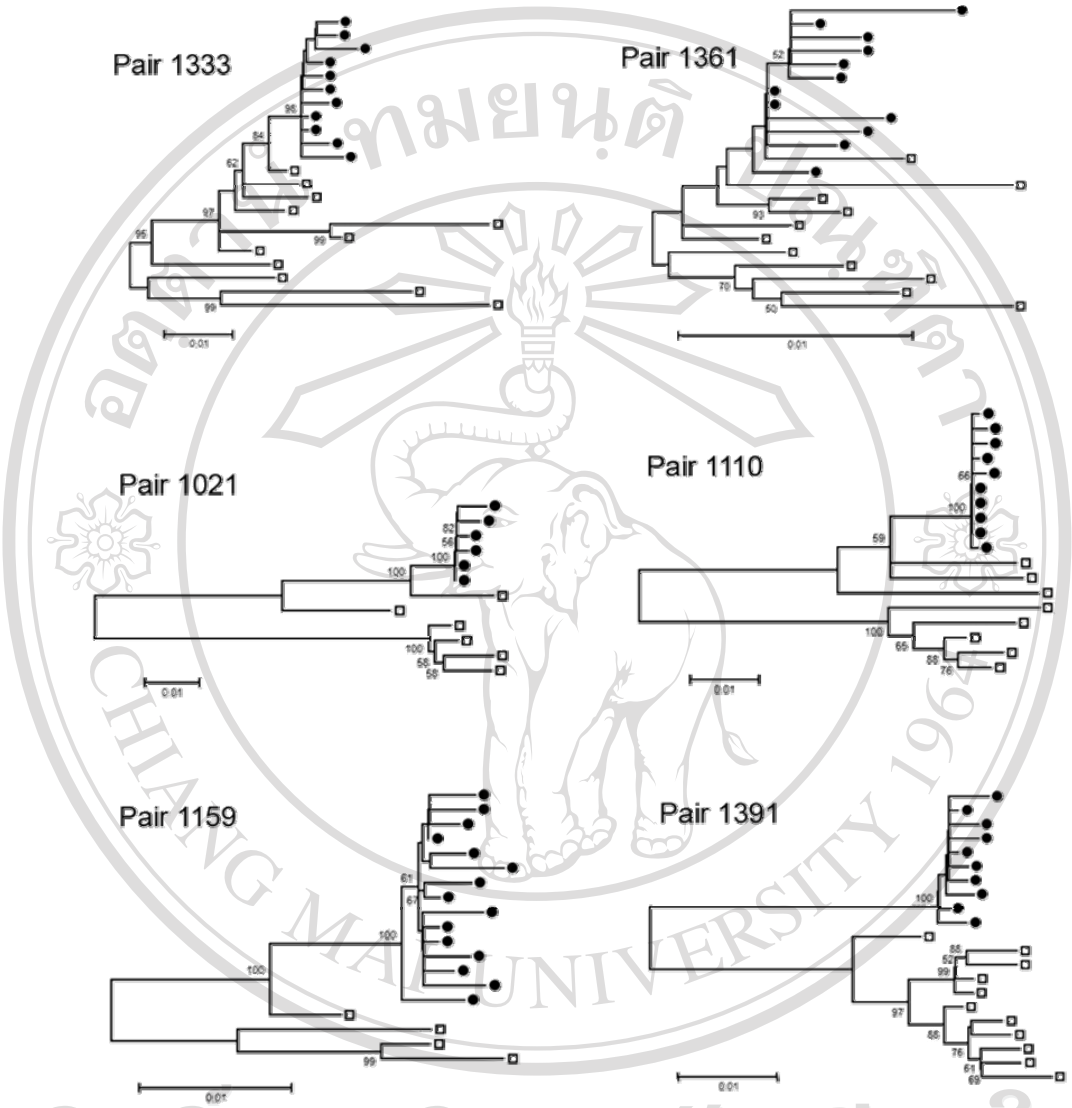


Figure 36. Unrooted neighbor-joining tree of the 353 HIV-1 env gp120 nucleotide sequences from 17 mother-child pairs. Horizontal branch lengths are drawn to scale, with the black bar representing 1% divergence. Bootstrap values are expressed as percentages per 1,000 replicates for each branch. Each symbol represents a single *env* sequence; squares, maternal sequences; circles, infant sequences. The symbols for each mother-infant pair are indicated with a different color. The number of each pair is indicated.

A.



A. (continued)



B.

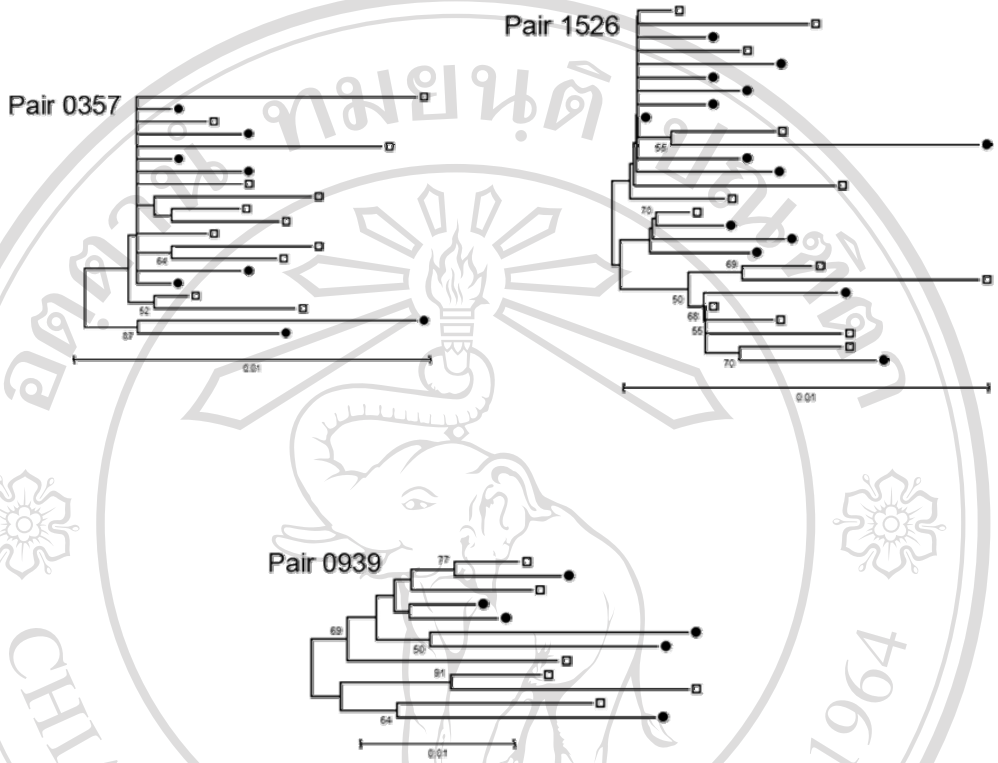


Figure 37. Unrooted neighbor-joining trees of HIV-1 *env* gp120 nucleotide sequences for each mother-child pair. (**A**) Trees of 14 pairs with transmission of a single maternal viral variant. (**B**) Trees of 3 pairs suggesting the transmission of several maternal variants. Bootstrap values are expressed as percentages per 1,000 replicates. Only bootstrap values higher than 50% are indicated. Horizontal branch lengths are drawn to scale, with the black bar representing 1% divergence. Each symbol represents a single *env* sequence; empty squares, maternal sequence; full circles, infant sequence.

The median intra-patient genetic distance was 2.4% (IQR 1.6 to 3.6%) in the mothers and 0.6% (IQR 0.6 to 0.8%) in the infants and this difference was significant overall ($P < 0.001$), as well as when the analysis was restricted to the cases of *in utero* transmission ($P = 0.05$) or to the cases of intrapartum transmission ($P < 0.01$) (Figure 38A). The distribution of intra-patient genetic distances was not significantly different in mothers who transmitted the virus *in utero* and those who transmitted intrapartum (2.5% and 2.4%, respectively; $P = 0.92$), or between infants infected *in utero* and infants infected intrapartum (median values: 0.7% and 0.6%, respectively; $P = 0.61$) (Figure 38A).

In fourteen pairs, sequences from the infants indicated the presence of a highly restricted homogeneous viral population, clustering in a single branch of the tree corresponding to a subcluster of the heterogeneous maternal viral population (Figure 37A). In the three remaining pairs, the viral env sequences obtained from the infants and their respective mothers were intermingled (Figure 37B). However, in two of these pairs, the intra-patient genetic distance was low in both mothers and infants (0.8% and 0.7% for the mother and infant of pair 0357, respectively, and 0.9% and 0.8% for the mother and infant of pair 1526, respectively). In only one pair (0939) was there clear evidence supporting the transmission of multiple divergent genetic variants from mother to infant. The intra-patient genetic distance was 3.0% and 2.9% in the mother and infant of this pair, respectively.

Lengths and potential N-linked glycosylation sites of the V1-to-V5 env sequences.

The median length of the V1-to-V5 *env* sequences from the mothers' viruses was 339 amino acids (IQR 336 to 344), not significantly different in the infants (median, 337; IQR 336 to 343) ($P = 0.64$) in the whole group of 17 pairs or according to the timing of transmission (Figure 38B). Similarly, we observed no difference when individual variables or constant regions were compared separately (data not shown).

There was a median number of 22 PNGS in the V1-to-V5 region (Fig. 38C), not significantly different between clones isolated from the mothers (IQR 21 to 24) and those from the infants (IQR 22 to 23) ($P = 0.83$), overall as well as when the analysis was restricted to the cases of *in utero* transmission ($P = 0.40$), and those of intrapartum transmission ($P = 0.31$).

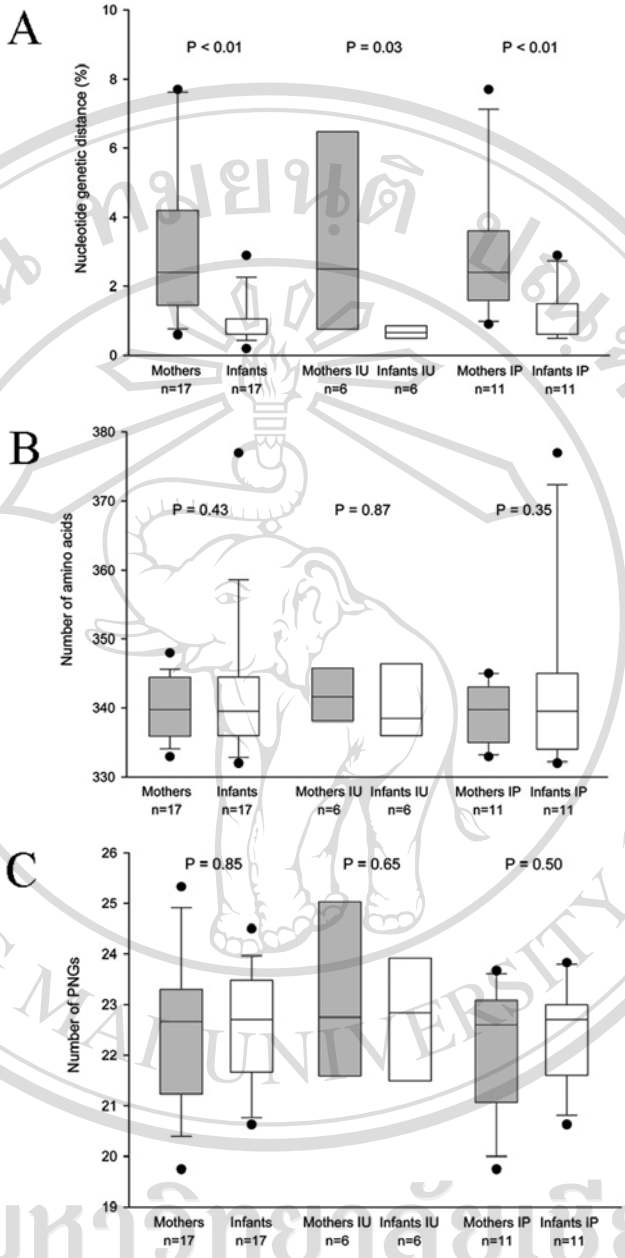


Figure 38. Characteristics of the viral variants in mothers and infants in all pairs (n=17), in cases of *in utero* transmission only (n=6), and in cases of intrapartum transmission only (n=11).

(**A**) Nucleotide intra-patient genetic distances

(**B**) Median V1-to-V5 lengths, as numbers of amino acids, in mothers and infants

(**C**) Median numbers of PNGS in mothers and infants,

P-value was calculated using a mixed-effects model (see Methods).

Median: horizontal bar in the box, 25th and 75th percentiles: lower and upper hinges, lower adjacent values: adjacent lines, outside value: points.

The numbers of PNGS were similar for viruses from mothers and infants, but some positions were highly conserved in all sequences, whereas others were more variable. We therefore tried to map the conserved positions in maternal viruses and in viruses from the infants (Figure 39). Eight PNGS, including N157 and N161 in the V1-V2 junction, N197, N262, N276, and N289 in C2, N333 in C3, and N440 in C4, were conserved in all viruses from mothers and infants, with a detection frequency of at least 80%, suggesting an important role in the maintenance of viral integrity. Conversely, a few PNGS were identified as signatures of isolates infecting a given pair, because they were found in all the clones from both mother and infant. These PNGS include for instance an additional site at the C-terminus of V1 in pair 779 and N434 in C4 of pair 0357. In a few cases, PNGS were found at a high frequency in unusual positions in a few infant variants only, and not in the corresponding maternal viral population, suggesting that these PNGS might confer an advantage to the transmitted virus. Examples of this situation are shown for pairs 1159 and 1224, in which > 90 % of the infant sequences harbor an additional PNGS at N397 in V4, and pairs 1005, 1110 and 1159, in which 100 % of the infant sequences harbor N454 in V5. A few PNGS were conserved in almost all infant virus sequences, whereas they were conserved, but present at a lower frequency in the maternal viruses. This was particularly marked for the PNGS at N301. This PNGS was present in 93 to 100% of clones from the infants of pairs 0858, 0978, 1021 and 1333, whereas it was totally absent from all 12 clones from mother 0858, and was present in only 21%, 67% and 46% of clones from mothers 0978, 1021 and 1333, respectively. Another PNGS, N384 in C3, displayed a similar pattern, except in pair 0938 in which it was found in only 25% of the infant sequences. The logistic regression model showed that the risk to find a PNGS was significantly higher (P<0.001) in the infant compared to the mother at the following four positions: 301, 397, 454, and 384, sorted by decreasing values of the odds ratio of the association. This suggests that the PNGS N301 and N384, shared by many CRF01_AE isolates, might confer a selective advantage on any transmitted virus of this clade.

gp120		V1								V1-V2		V2				C2								V3				C3						V4						C4			V5		
position		129	+	134	137	140	143	150	+	157	161	+	+	187	188	189	197	234	241	262	276	289	295	301	305	316	320	333	338	354	356	384	390	396	397	403	+	+	434	440	443	453	454	456
0357	MO	0	0	100	0	100	0	100	0	91.7	100	0	0	0	100	0	100	100	100	100	100	100	100	100	0	0	0	100	0	100	0	83.3	100	100	0	100	0	0	100	100	0	100	0	0
0377	MO	100	0	50	66.7	50	50	100	0	100	100	0	0	0	33.3	0	100	100	100	100	100	100	66.7	91.7	0	0	0	100	100	100	0	58.3	91.7	66.7	0	66.7	66.7	25	0	100	0	100	0	100
0540	MO	0	0	91.7	0	16.7	41.7	100	0	100	100	0	0	83.3	91.7	0	100	100	100	100	100	100	100	100	0	0	0	100	0	100	0	100	100	100	0	100	0	0	0	100	0	100	0	100
0779	MO	73.3	6.7	40	0	20	73.3	100	100	100	100	0	0	0	93.3	0	100	100	100	100	100	93.3	40	100	0	0	0	93.3	20	93.3	0	100	100	93.3	0	53.3	0	0	0	100	0	86.7	0	100
0858	MO	0	0	0	0	0	91.7	100	0	100	100	0	0	0	100	0	100	100	100	100	100	100	100	0	0	0	0	100	0	100	0	0	100	100	0	100	100	0	0	100	0	92.9	0	83.3
0938	MO	0	0	75	0	100	0	100	0	100	100	0	0	0	100	0	100	0	100	81.8	100	100	100	100	0	0	0	100	0	100	25	12.5	100	100	0	100	0	0	0	100	0	100	0	0
0939	MO	50	0	33.3	0	100	100	100	0	83.3	83.3	0	0	0	100	0	100	100	100	100	100	100	100	100	0	0	0	83.3	0	83.3	0	100	100	100	0	100	0	0	0	100	0	100	0	100
0978	MO	0	0	100	0	100	0	92.9	0	100	100	0	0	0	100	0	100	100	100	100	100	100	100	21.4	0	0	0	100	100	100	0	100	21.4	71.4	0	100	0	0	0	100	0	100	0	0
1005	MO	100	0	100	0	100	50	50	0	100	100	50	100	0	100	0	100	83.3	16.7	100	100	100	100	100	0	0	0	83.3	83.3	100	0	100	83.3	100	0	100	0	66.7	0	100	0	66.7	16.7	100
1021	MO	0	0	100	0	100	33.3	33.3	0	100	100	0	0	0	100	0	100	100	100	100	83.3	100	66.7	66.7	0	0	0	83.3	0	100	0	33.3	100	83.3	0	83.3	0	66.7	0	100	0	100	0	100
1110	MO	87.5	0	75	0	87.5	75	100	0	87.5	87.5	0	0	0	100	0	100	87.5	100	100	100	100	100	100	0	0	0	100	0	100	0	100	87.5	87.5	0	75	0	0	0	100	0	62.5	37.5	62.5
1159	MO	0	0	100	0	100	0	100	0	100	100	0	0	0	100	50	100	100	100	100	100	100	100	100	0	0	0	100	0	75	0	100	100	75	25	100	0	0	0	100	0	25	25	50
1224	MO	100	0	93.3	0	100	100	100	0	100	100	0	0	0	100	0	100	100	100	100	86.7	100	100	100	0	6.7	0	100	0	100	33.3	100	60	100	20	80	0	0	0	100	0	93.3	13.3	100
1333	MO	27.3	0	18.2	0	18.2	81.8	90.9	0	100	100	0	0	0	100	0	100	100	100	90.9	100	100	100	45.5	0	0	0	81.8	0	100	0	90.9	100	90.9	0	90.9	0	0	0	81.8	0	100	0	100
1361	MO	90.9	0	0	0	90.9	0	100	0	100	100	0	0	100	100	0	100	18.2	100	100	100	100	100	100	0	0	0	100	0	72.8	0	100	100	100	0	100	0	0	0	90.9	9.1	72.8	0	90.9
1391	MO	45.5	0	63.6	0	100	54.5	54.5	0	100	100	0	0	0	100	0	100	100	100	90.9	100	100	100	100	0	0	0	90.9	100	72.7	0	100	100	0	0	100	0	0	0	100	0	0	0	100
1526	MO	100	0	100	0	100	38.5	100	0	100	100	0	0	0	100	0	100	100	100	100	100	100	100	100	0	0	0	100	0	100	0	100	100	0	0	100	0	0	0	100	0	100	0	100
0357	IN	0	0	100	0	100	0	100	0	100	100	0	0	0	87.5	0	100	100	100	100	100	100	100	100	0	0	0	87.5	0	87.5	0	100	100	87.5	0	100	0	0	100	100	0	87.5	0	0
0377	IN	100	0	100	0	0	0	90.9	0	100	100	0	0	0	0	0	100	100	100	100	90.9	100	90.9	90.9	0	0	0	100	100	90.9	0	100	100	100	0	0	0	100	0	100	0	0	0	100
0540	IN	0	0	92.9	0	0	0	100	0	100	100	0	0	100	100	0	100	100	100	92.9	92.9	92.9	100	100	0	0	0	100	0	100	0	100	100	100	0	100	0	0	0	100	0	100	0	100
0779	IN	0	0	0	0	42.9	0	100	100	100	100	0	0	0	100	0	100	100	100	100	100	100	100	100	0	0	0	100	100	100	0	100	100	100	0	0	0	0	0	100	0	100	0	100
0858	IN	0	0	0	0	0	92.9	100	0	100	100	92.9	92.9	0	92.9	0	100	100	92.9	100	92.9	100	100	92.9	0	0	0	100	0	100	0	92.9	100	100	0	92.9	0	0	0	100	0	100	0	85.7
0938	IN	0	0	100	0	100	0	100	0	100	100	0	0	0	100	0	100	0	100	100	100	100	100	100	0	0	0	100	0	100	0	25	100	87.5	0	100	0	0	0	100	0	100	0	25
0939	IN	83.3	0	83.3	0	100	100	100	0	100	83.3	0	0	0	100	0	100	83.3	100	83.3	100	100	100	100	0	0	0	100	0	100	0	100	83.3	100	0	100	0	0	0	100	0	100	0	100
0978	IN	0	0	100	0	93.3	0	93.3	0	100	100	0	0	0	100	0	100	100	100	100	100	100	100	100	0	0	0	100	100	100	0	100	0	100	0	93.3	0	0	0	100	0	80	0	0
1005	IN	100	0	87.5	0	100	100	0	0	100	100	0	0	0	100	0	100	100	100	100	100	100	100	100	0	0	0	100	100	100	0	87.5	25	100	0	100	0	100	0	100	0	50	100	0
1021	IN	0	0	100	0	100	100	33.3	0	100	100	0	0	0	100	0	100	100	100	100	100	100	0	100	0	0	0	100	0	100	0	100	100	100	0	100	0	0	0	100	0	100	0	0
1110	IN	100	0	0	0	100	100	100	0	100	100	0	0	0	100	0	100	100	100	100	100	100	100	100	0	0	0	100	0	100	0	90	100	0	0	0	0	0	0	100	0	0	100	0
1159	IN	100	0	0	0	93.3	0	100	0	100	100	0	0	0	100	0	100	100	100	93.3	100	100	100	100	0	0	0	100	0	100	0	100	93.3	0	100	100	0	0	0	93.3	0	0	100	0
1224	IN	100	0	0	0	100	100	100	0	100	100	0	0	0	100	0	93.3	93.3	100	100	100	100	100	100	0	0	0	90.9	0	100	0	100	0	90	90	100	0	0	0	100	0	100	0	100
1333	IN	0	0	18.2	0	0	100	81.8	0	100	100	0	0	0	100	0	100	100	100	100	100	100	100	100	0	0	0	100	0	90.9	0	100	100	100	0	90.9	0	0	0	100	0	100	0	100
1361	IN	100	0	0	0	100	0	100	0	100	91.7	0	0	100	91.7	0	100	100	91.7	100	100	100	100	100	0	0	8.3	91.7	0	100	0	100	100	100	0	91.7	0	0	0	100	8.3	100	0	100
1391	IN	100	0	100	0	100	0	100	0	100	100	0	0	0	100	0	100	100	100	100	100	100	100	100	0	0	0	100	0	100	0	100	100	100	0	100	0	0	0	100	0	100	0	0
1526	IN	100	0	100	0	100	14.3	100	0	92.9	100	0	0	0	100	0	100	92.9	100	100	100	100	100	7.1	0	0	0	100	0	100	0	100	85.7	0	0	100	0	0	0	100	0	100	0	100

Figure 39. Map of PNGS for each quasispecies present in either mother or infant. We recorded the positions of PNGS in the V1-to-V5 sequences from each of the 353 clones, with respect to the CRF01_AE consensus sequence. We then calculated the frequency of each PNGS in the sequences of each mother or infant. PNGS were considered to be conserved at a given position if they were present in more than 80% of the clones from a given individual The positions of the PNGS correspond to those of the consensus HIV-1 CRF01_AE obtained from the Los Alamos National Laboratory (http://hiv.lanl.gov). Additional PNGS found in our subjects are indicated with the symbol (+). The frequency of PNGS at each position is presented as a percentage for each case. The mothers are indicated by an ID number followed by MO, and the infants by an ID number followed by IN. Values greater than 80% are highlighted in black, values between 50 and 79.9% are highlighted in gray, and those lower than 50% are highlighted in white.

4. Discussion

We analyzed 353 sequences encompassing almost the entire gp120 *env* gene from 17 mother-infant pairs infected with HIV-1 CRF01_AE strains, to investigate the characteristics of the envelope genes involved in the MTCT of HIV-1. Several previous studies with similar aims have been carried out [428; 357; 301; 410; 109; 432; 111; 223; 338; 398], but our study is probably one of the most extensive and homogeneous, due to the limited genetic diversity of the viruses circulating in Thailand. All the sets of sequences from infants clustered closely with the corresponding maternal sequences, but could be distinguished from the maternal sequences in 14 of the 17 pairs, suggesting a significant bottleneck at transmission. In these 14 infants, viral sequences were highly homogeneous, with a much shorter genetic distance than that of the viruses in the corresponding mothers. Multiple variants were clearly shown to be transmitted from mother to infant in only one pair, with similar large genetic distances between the infant and maternal variants. Phylogenetic analysis of the variants from the *env* sequences from the mothers and infants of the two remaining pairs showed these sequences to be intermingled. However, there was a limited genetic distance between the maternal sequences, suggesting that the evolution of the virus was restricted in these mothers. It is therefore unclear whether these two cases can really be classified as multiple variant transmission. We found no difference in genetic diversity between viruses present in infants infected *in utero* and viruses present in infants infected intrapartum, though, due to the small sample size, the power to detect such differences was only 25% in mothers and 16% in infants. This result is consistent with the findings of some [410; 111; 338] but in not all [111] previous studies. Indeed, Dickover *et al*. suggested that *in utero* transmitters were more likely to transmit multiple variants than intrapartum transmitters [111]. Our results confirm the transmission of a genetically restricted viral population during MTCT, with no difference as a function of the timing of transmission.

Extensive diversity was observed in the mother envelope sequences whereas the majority of infant samples showed limited viral diversity. We believe that this difference was not due to re-sampling bias during the PCR steps, particularly because there was a high plasma VL in most of the infants and we used the entire amount of extracted RNA for reverse transcription. Therefore this suggests that there was no limitation in number of viral templates during the 1st round of PCR. In addition, we performed subsequent multiple independent PCRs for several of these infant samples, that were pooled before cloning and sequencing. A similar limited viral diversity was observed. Finally, additional studies performed on the contemporary infant PBMCs showed that the same limited viral diversity was found in both PBMC samples and plasma samples (data not shown).

Studies of viruses isolated in early infection are of prime importance because these viruses correspond to variants capable of successfully establishing infection in natural conditions. Descriptions of the characteristics of these early viruses are particularly important for the development and rational design of efficient vaccines. Recent studies have highlighted potential differences in the biology of sexual

transmission between viral subtypes [107; 79; 132; 44]. Viruses of subtypes A and C appear to pass through a genetic bottleneck during or shortly after heterosexual transmission, leading to the selection of viruses with a compact and less glycosylated gp120 [107; 79]. This property was not observed for sexually transmitted subtype B viruses [79; 132]. Descriptions of the molecular properties of vertically transmitted viruses in infants, at the time at which the first positive test result is obtained, provides additional information because these viruses are successfully transmitted to infants despite the presence of pre-existing passively transferred antibodies, particularly if transmission occurs intrapartum or late in pregnancy [432]. They therefore provide excellent examples of neutralization escape variants potentially able to improve our understanding, for the benefit of future vaccine approaches. Based on recent observations suggesting that sexually transmitted viruses have shorter gp120 proteins and fewer PNGS, we examined these sequence features in the gp120 genes of our mother-child pairs. We observed no difference between the viruses of the mothers and infants, in terms of both the length of the V1-to-V5 region and number of PNGS (even if each variable loop was considered separately). These findings differ slightly from those of the recent study by Wu *et al*., in which the V1-to-V5 lengths of the envelope did not differ between infants and mothers, but fewer PNGS were found in viruses from infants than in those from their mothers infected with viruses of various subtypes, A, C, or C/D and D/A recombinants [432]. Although our study focused on a non-breastfeeding group whereas that of Wu *et al*. looked at a breastfeeding population, this discrepancy may be due also to differences in the biology of different subtypes, and highlights the need to carry out such studies in various continents, with viruses from diverse clades.

The numbers of PNGS were similar in the viruses from mothers and infants in our study, but we nonetheless analyzed their location in gp120. The dense carbohydrate (glycan) array surrounding the HIV envelope protein is now known to evolve continually during natural infection— a powerful adaptation for escape from neutralizing antibodies and virus survival in the host [99; 355; 420]. We therefore mapped the PNGS that were conserved or variable in all mother-infant pairs. The most striking observation was the high degree of conservation of two PNGS in infant viruses, at positions N301 in V3 and N384 in C3. These PNGS were significantly less conserved in maternal viruses ($P<0.001$ in a logistic regression model). This pattern suggests that these PNGS may be associated with a selective advantage, at least for CRF01_AE viruses. This could be confirmed using, for example, pseudotyped viruses, and investigating the possible association of these PNGS with better fitness and/or more efficient escape from neutralization. The PNGS at position N301 was present in 0 to 67% of the sequences from four mothers, whereas it was detected in almost all the clones from the corresponding infants. Interestingly enough, this N301 glycan site was also found in all maternally transmitted CRF01_AE viruses sequenced in an independent study in Thailand [384]. In addition, several *in vitro* studies have shown that N301 is associated with both a decrease in the sensitivity of HIV-1 to neutralization by CD4BS antibodies and modulation of the interaction of the HIV-1 envelope with CD4 and chemokine receptors [199; 205; 247]. The removal of carbohydrate at position 301 could involve rearrangements within either the same monomer or the adjacent monomer. Koch *et al*. demonstrated that the deletion of

N301 had no direct effect on CD4BS Abs recognition in the context of monomeric gp120, but it allowed a repositioning of the V1/V2 stem on the adjacent monomer or some other element of gp120, which became less effective at shielding the CD4 binding site [205].

In conclusion, we analyzed the molecular characteristics of transmitted HIV-1 viruses of the CRF01_AE clade in MTCT. We also provide evidence that, despite the presence of a complex viral population in the mother, only a restricted subset of viruses is transmitted to the infant, regardless of whether transmission occurs *in utero* or intrapartum. We did not find that more compact (shorter) gp120s or fewer PNGS were characteristic of viruses transmitted from mother to infant, as suggested for at least a few subtypes of sexually transmitted viruses. However, our data suggest that a limited number of PNGS that appear to be conserved in all infant variants may confer an advantage on the virus to be transmitted, probably allowing it to escape antibody-mediated neutralization, as suggested by the most recent studies in the field of MTCT [109; 432]. Functional studies are required to confirm these findings, but our results nonetheless highlight the role of the "glycan shield" in HIV biology, and immune escape in particular [409]. Dissection of the molecular characteristics of viruses transmitted to infants in MTCT, in which exposure to the virus occurs in the presence of passively transferred pre-existing antibodies, could provide key elements for the prevention of HIV infection through vaccination or passive immunization.

5. Publications

This work has been accepted for publication in the Journal of Infectious Diseases (September 2008, vol. 198(6): pages 868-876.).

> **Samleerat T**, Braibant M, Jourdain G, Moreau A, Ngo-Giang-Huong N, Leechanachai P, Hemvuttiphan J, Hinjiranandana T, Changchit T, Warachit B, Suraseranivong V, Lallemant V, Barin F. *Characteristics of HIV-1 glycoprotein 120 env sequences in mother-infant pairs infected with HIV-1 subtype CRF01_AE*. J Infect Dis 2008; 198(6):868-876.

This work was presented in part at The AIDS Vaccine 2007 Conference. Seattle, Washington USA. August 20-23, 2007, and at the 4th Dominique Dormont International Conference. Paris, France. December 13-15, 2007.

> **Samleerat T**, Braibant M, Jourdain G, Moreau A, Ngo-Giang-Huong N, Leechanachai P, Hemvuttiphan J, Hinjiranandana T, Changchit T, Warachit B, Suraseranivong V, Lallemant V, Barin F. *Molecular characteristics of the HIV-1 envelope glycoproteins of CRF01_AE variants transmitted from mother to child*. The AIDS Vaccine 2007 Conference. Seattle, Washington USA. August 20-23, 2007 ***(oral presentation-presenting author)***.
>
> Abstract published in: Antiviral Therapy 2007; Suppl 4: 1-210

Samleerat T, Braibant M, Jourdain G, Moreau A, Ngo-Giang-Huong N, Leechanachai P, Hemvuttiphan J, Hinjiranandana T, Changchit T, Warachit B, Suraseranivong V, Lallemant V, Barin F. *Characteristics of HIV-1 gp120 env sequences in mother-child pairs infected with HIV-1 subtype CRF01_AE*. 4th Dominique Dormont International Conference. Maternal chronic viral infections transmitted to infants: from mechanisms to prevention and care. Paris, France. December 13-15, 2007 ***(poster presentation-presenting author)***.

Abstract published in: Retrovirology 2008, 5(Suppl 1):P5

MAJOR ARTICLE

Characteristics of HIV Type 1 (HIV-1) Glycoprotein 120 *env* Sequences in Mother-Infant Pairs Infected with HIV-1 Subtype CRF01_AE

Tanawan Samleerat,[1,9] Martine Braibant,[9] Gonzague Jourdain,[2,3] Alain Moreau,[9] Nicole Ngo-Giang-Huong,[2,3] Pranee Leechanachai,[1] Jittapol Hemvuttiphan,[4] Temsiri Hinjiranandana,[5] Tikamporn Changchit,[6] Boonyarat Warachit,[7] Veera Suraseranivong,[8] Marc Lallemant,[2,3] and Francis Barin[9]

[1]Faculty of Associated Medical Sciences, Chiang Mai University, [2]Institut de Recherche pour le Développement, UMI 174, and [3]Harvard School of Public Health, Chiang Mai, [4]Phayao Hospital, Phayao, [5]Somdej Pranangchao Sirikit Hospital, Chonburi, [6]Phan Hospital, Chiang Rai, [7]Hat Yai Hospital, Songkla, and [8]Bhumibol Adulyadej Hospital, Bangkok, Thailand; [9]Université François-Rabelais, Institut National de la Santé et de la Recherche Medicale ERI 19 and Centre Hospitalier Régional Universitaire de Tours, Tours, France

We analyzed the characteristics of the envelope genes of human immunodeficiency virus type 1 in 17 mother-infant pairs infected with variants of the CRF01_AE clade. A total of 353 sequences covering almost the entire glycoprotein (gp) 120 region were available for analysis. We found that, even if the virus population in the mother was complex, only viruses of a restricted subset were transmitted to her infant, independently of whether transmission occurred in utero or during the intrapartum period. We did not find that shorter gp120 regions or fewer potential N-glycosylation sites (PNGS) were characteristic of viruses transmitted from mother to infant. However, our data suggest that a limited number of PNGS that seem to be conserved in all variants in infants but are not uniformly present in variants in mothers may confer an advantage for transmission of the virus, thereby highlighting the potentially important role of the "glycan shield." This finding was particularly significant for the PNGS at positions N301 and N384.

Mother-to-child transmission (MTCT) of HIV type 1 (HIV-1) can occur during pregnancy (in utero), during labor and delivery (the intrapartum period), or after birth (through breast-feeding) [1, 2]. The risk of MTCT can be reduced to <2% by use of such interventions as antiretroviral prophylaxis, elective cesarean delivery, and complete avoidance of breast-feeding [3–5].

HIV-1 evolves continually during the course of infection in an individual [6, 7]. Evolution within the host is driven by continuous selection pressure resulting from host immune responses [8, 9]. The *env* gene is the HIV-1 gene with the highest rate of evolution, and the HIV-1 envelope glycoproteins—in particular, gp120—are the targets of neutralizing antibodies [10, 11].

Previous studies have demonstrated that HIV-1–infected mothers generally have a heterogeneous virus population, whereas their infected infants have a more homogeneous virus population [12–16]. Maternal antibodies are among the selective factors that are potentially responsible for this genetic bottleneck. Indeed, several studies have shown that mothers who do not transmit HIV-1 are more likely to have higher levels of neutralizing antibodies than are mothers who do transmit HIV-1, a finding that is consistent with these antibodies having a role in reducing MTCT [17–21]. Supporting this hypothesis, recent studies have clearly shown that variants in infants are frequently resistant to neutralization by autologous maternal plasma [22, 23]. However, infants occasionally have been found to be infected with a heterogeneous population of multiple maternal variants [13, 15, 16, 24–26]. It has been suggested that the timing of transmission is responsible for differ-

Received 14 December 2007; accepted 24 March 2008; electronically published 22 August 2008.

Potential conflicts of interest: none reported.

Presented in part: AIDS Vaccine Conference, Seattle, Oregon, 20–23 August, 2007.

Financial support: Agence Nationale de Recherches sur le Sida (Paris, France); Institut de Recherche pour le Développement (Paris, France); National Institutes of Health (grant 5 R01 HD 33326); Ministry of Foreign Affairs (France) and the Commission of Higher Education, Ministry of Education (Thailand) (doctoral fellowships to T.S.); Sidaction (Paris, France) (postdoctoral fellowship to M.B.).

Reprints or correspondence: Dr. Francis Barin, Laboratoire de Virologie, CHU Bretonneau, 2 blvd. Tonnellé, 37044 Tours cedex, France (fbarin@med.univ-tours.fr).

The Journal of Infectious Diseases 2008; 198:868–76

0022-1899/2008/19806-0012$15.00
DOI: 10.1086/591251

ences in the diversity of the virus population in infants, with in utero transmitters more likely to transmit single or multiple maternal variants and with intrapartum transmitters more likely to transmit minor HIV-1 variants [24]. However, the existence of different selective pressures, depending on the timing of transmission, was not confirmed in other studies [16, 22].

The most extensive molecular studies, which included a maximum of 12–14 mother-infant pairs, were conducted in populations infected with different subtypes and involved the use of different molecular approaches [16, 22, 23]. In most studies in which the *env* gene was sequenced, only portions of the gp120 region were analyzed, with only one study comparing the entire V1–V5 region of 78 clones from 12 pairs [23].

We investigated the molecular characteristics of the HIV-1 variants involved in MTCT, by focusing on a homogeneous population infected with CRF01_AE isolates in Thailand and by analyzing entire gp120 *env* sequences (353 clones) from 17 mother-infant pairs for whom the timing of transmission was known. Because recent studies of heterosexually transmitted viruses identified soon after infection have suggested that these viruses have fewer potential N-linked glycosylation sites (PNGS) and shorter variable loop sequences in the envelope than do transmitted isolates identified later after infection, depending on subtype [27–30], we examined these sequence characteristics, to look for molecular characteristics that might confer a selective advantage for MTCT.

MATERIALS AND METHODS

Study population and sample collection. Samples were obtained from participants enrolled in a clinical trial for the prevention of MTCT in Thailand [31]. The infants were not breast-fed. The HIV-1 infection status of the infants was determined by means of polymerase chain reaction (PCR)–based detection of HIV-1 proviral DNA, with the use of DNA extracted from peripheral blood samples collected within 48 h of birth, at 6 weeks after birth, and at 4 and 6 months of age. Infants were considered to be uninfected if the results of 2 PCR assays for the detection of HIV DNA were negative on 2 separate occasions after 1 month of age. Infants were considered to be HIV infected if the test result was positive on 2 separate occasions. We considered the infant to have been infected in utero if the first test result, obtained for blood collected within 48 h of birth, was positive, and to have been infected in the intrapartum period if the first test result was negative but subsequent test results were positive [32]. We randomly selected 24 HIV-1–infected mother-infant pairs for whom samples were available. We analyzed maternal peripheral blood samples obtained from the mother at the time of delivery and plasma samples obtained from the infant either at or as close as possible to the first time point at which the HIV-1 DNA PCR result was positive. The median time for obtaining plasma samples from infants was 44 days after delivery (interquartile range [IQR], 39–49 days after delivery). The median plasma viral load in infants was 4.7 $\log_{10}$ copies (range, 2.2–5.1 $\log_{10}$ copies). We focused on viral sequences corresponding to the proviral DNA present in the peripheral blood mononuclear cells (PBMCs) of the mothers, rather than the RNA present in their plasma virions, to obtain more-complete information about the evolution of the virus population during infection. Furthermore, the chemoprophylaxis administered to the mothers frequently decreased the viral load to levels too low to allow for amplification of the entire gp120 *env* fragment from viral RNA. In contrast, we focused on RNA corresponding to plasma virions from the infants, for description of the genetic characteristics of the early virus population that was successfully transmitted. All experiments were performed with investigators blind to the timing of transmission.

Amplification and cloning of HIV-1* env *genes. Genomic DNA was extracted from maternal peripheral blood by use of the QIAamp DNA Blood Midi Kit (Qiagen). Viral RNA was extracted from infant plasma samples (280 μL) by use of the QIAamp Viral RNA Mini Kit (Qiagen), and cDNA was synthesized from the entire extracted RNA with use of the SuperScript First-Strand Synthesis System for reverse-transcriptase (RT)–PCR (Invitrogen), with use of the specific primer AEa1ext (5'-TTCTCTTTGCCYTGGTGGGTGCTA-3'). A 1.2-kb fragment of the V1–V5 region covering almost the entire HIV-1 *env* gp120 gene (from upstream V1 to downstream V5) was amplified using nested PCR (Platinum PCR SuperMix High Fidelity; Invitrogen) with the use of subtype-specific primers. The primers AEs1ext (5'-TGGGTTACAGTTTATTATGGGG-3') and AEa1ext were used in the first round of amplification, and the primers AEs1int (5'-TGCCAAAGCATATGAGACAGARGYGCA-3') and AEa1int (5'-TTCACTTCTCCAATTGTCCTTYATRTT-3') were used in the second round. Plasma virus-derived cDNA (2 μL corresponding to 10% of the first strand reaction) or proviral DNA (1 μg) was used as a template for the first-round PCR and was amplified under the following conditions: heating at 94°C for 2 min; 35 cycles at 94°C for 30 s, at 50°C for 30 s, and at 68°C for 90 s; and a final extension step at 72°C for 10 min. The first PCR product was purified using the NucleoSpin Extract II Kit (Macherey-Nagel), to adjust the total volume of the PCR product to 10 μL for use as a template for the second round of PCR, which was performed under the same cycling conditions. PCR products were obtained from both the mother and the infant in 17 pairs. We therefore focused our investigation on these 17 pairs.

The 1.2-kb PCR products were gel purified using the NucleoSpin Extract II Kit and were inserted into the pCR2.1-TOPO TA cloning vector (Invitrogen). Individual bacterial colonies were screened for the presence of inserted *env* gene fragments by means of restriction enzyme analysis of the plasmid DNA.

Sequencing and genetic analysis. A total of 8–15 positive clones were sequenced for each subject, by use of 6 overlapping primers encompassing the entire gp120 sequence. Nucleotide

sequences were assembled and translated into amino acid sequences with use of the BioEdit package [33]. Nucleotide and deduced amino-acid sequences were aligned with the Clustal W program [34] and were manually adjusted to optimize the alignment. Sequences with premature stop codons were excluded. The HIV-1 subtype references were obtained from the HIV databases at Los Alamos National Laboratory [35] and were used to determine the HIV-1 subtype of all sequences from the 17 mother-infant pairs. Neighbor-joining trees were constructed using nucleotide sequences spanning the V1–V5 region and MEGA software (version 3.1) [36]. Branching-order reliability was assessed using bootstrap analysis with 1000 replicates. PNGS were identified using N-GlycoSite [37]. Genetic distances were calculated using the Kimura 2-parameter method (transition-to-transversion ratio, 2.0) [38].

Statistical analysis. We used the Wilcoxon signed-rank test to compare the intrapatient genetic distances between the mother and infant sequences, both overall and according to the timing of transmission, and we also used the Kruskal-Wallis test to compare cases of in utero and intrapartum transmission among mothers or infants. To study the differences between the length and the number of PNGS of the mothers' and infants' sequences, taking into account their distribution in the mother and in the infant, we used a linear mixed-effects model fit by maximum restricted likelihood, where mother-infant pairs were considered to be a random effect and where the variable "mother or infant" was considered to be both a fixed and a random effect.

In an exploratory analysis, we fitted an exact logistic regression model conditional on the number of PNGS within pairs (Procedure Exlogistic; Stata software [version 10; Stata]) to estimate, for each position, the ratio of the odds of the presence of a PNGS in infants' clones to the odds of a PNGS in mothers' clones, as well as to rank the amino acid positions in the gp120 accordingly. To take into account the multiple comparisons, we considered only those differences that were significant at $P \leq .001$.

Nucleotide sequence accession numbers. All sequences (353 V1–V5 *env* fragments) were deposited in GenBank (accession numbers EU031053–EU031405).

RESULTS

Characteristics of the patients. Among the 17 pairs, 6 infants were considered to have been infected in utero, and 11 were considered to have been infected during the intrapartum period.

Phylogenetic analysis of HIV-1* env *sequences from mother-infant pairs. We analyzed 353 envelope clones (175 clones from mothers and 178 clones from infants). All 353 V1–V5 sequences clustered with CRF01_AE reference sequences, with a high bootstrap value (data not shown), indicating that all cases were infected with viruses of the HIV-1 CRF01_AE clade. The neighbor-joining tree showed that the sequences from each pair formed a monophyletic group that was separated from the sequences of the other mother-infant pairs, indicating an epidemiological link between each mother and her infant and an absence of cross-sample contamination (figure 1). The interpair genetic distance (median, 14.5%; IQR, 12.8%–16.3%) was significantly higher ($P < .01$) than the intrapair genetic distance (median, 3.1%; IQR, 2.1%–5.8%) (figure 2).

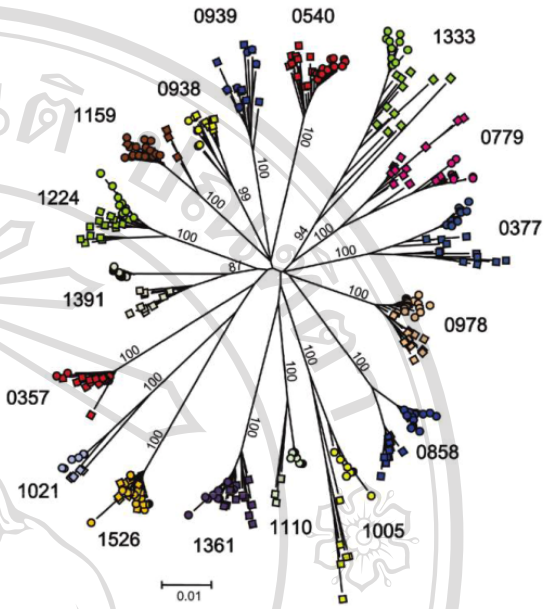


Figure 1. Unrooted neighbor-joining tree of the 353 HIV-1 *env* gp120 nucleotide sequences from 17 mother-infant pairs. Horizontal branch lengths are drawn to scale, with the black bar denoting 1% divergence. Bootstrap values are expressed as percentages per 1000 replicates for each branch. Each symbol denotes a single *env* sequence; □, maternal sequence; ●, infant sequence. The symbols for each mother-infant pair are presented in a different color. The number of each pair is indicated.

The median intrapatient genetic distance was 2.4% (IQR, 1.6%–3.6%) in the mothers and 0.6% (IQR, 0.6%–0.8%) in the infants, and this difference was significant overall ($P < .001$), as well as when the analysis was restricted to cases of in utero transmission ($P = .05$) or to cases of intrapartum transmission ($P < .01$) (figure 3*A*). The distribution of intrapatient genetic distances was not significantly different between mothers who transmitted the virus in utero and those who transmitted the virus during the intrapartum period (2.5% and 2.4%, respectively) ($P = .92$) or between infants infected in utero and infants infected during the intrapartum period (median values, 0.7% and 0.6%, respectively) ($P = .61$) (figure 3*A*).

In 14 pairs, sequences from the infants indicated the presence of a highly restricted homogeneous virus population, clustering in a single branch of the tree corresponding to a subcluster of the heterogeneous maternal virus population (figure 2A). In the 3 remaining pairs, the viral *env* sequences obtained from the infants and their

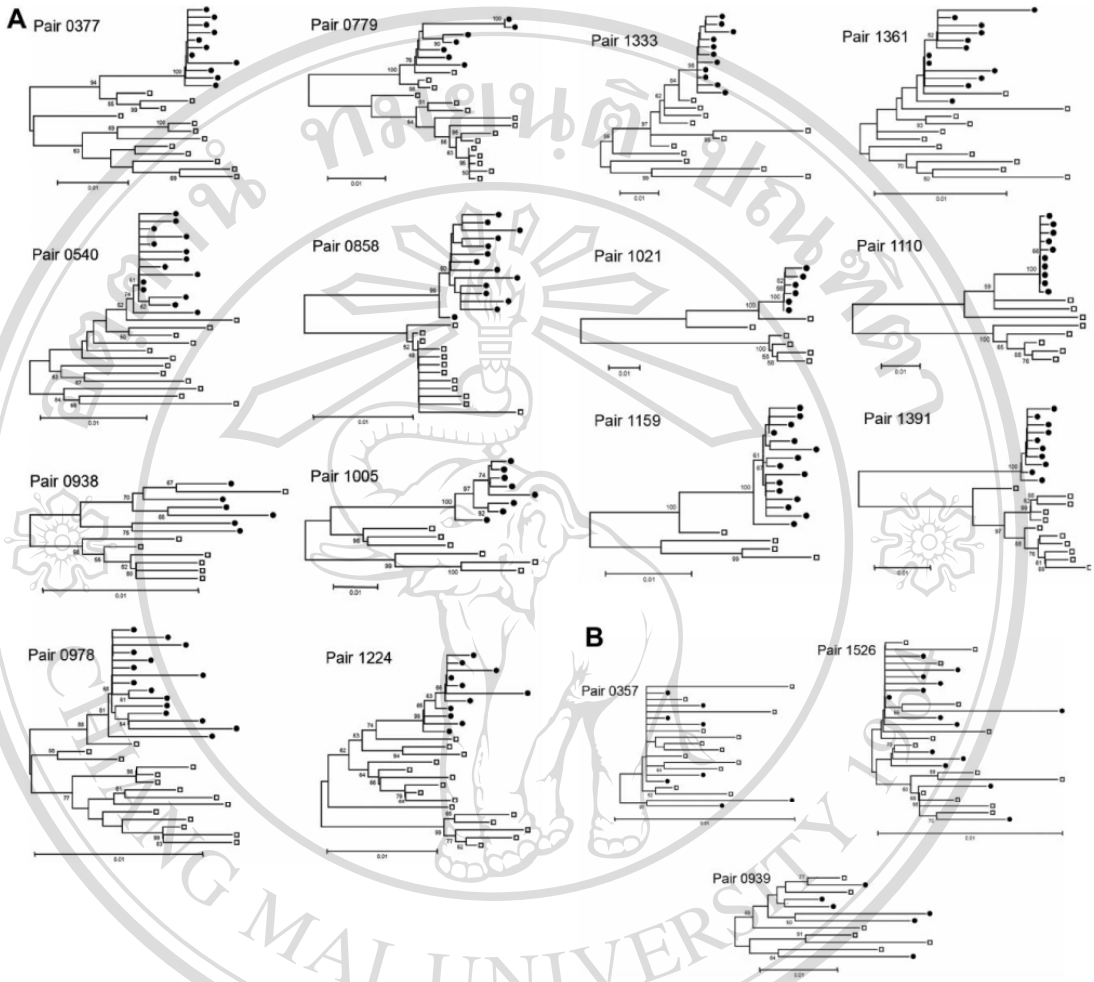


Figure 2. Unrooted neighbor-joining trees of HIV-1 *env* gp120 nucleotide sequences for each mother-infant pair. *A,* Trees of 14 pairs with transmission of a single maternal viral variant. *B,* Trees of 3 pairs suggesting the transmission of several maternal variants. Bootstrap values are expressed as percentages per 1000 replicates. Only bootstrap values >50% are indicated. Horizontal branch lengths are drawn to scale, with the black bar denoting 1% divergence. Each symbol denotes a single *env* sequence; □, maternal sequence; ●, infant sequence.

respective mothers were intermingled (figure 2*B*). However, in 2 of these pairs, the intrapatient genetic distance was low in both mothers and infants (0.8% and 0.7% for the mother and the infant of pair 0357, respectively, and 0.9% and 0.8% for the mother and the infant of pair 1526, respectively). In only one pair (pair 0939) was there clear evidence supporting the transmission of multiple divergent genetic variants from mother to infant. The intrapatient genetic distance was 3.0% and 2.9% in the mother and the infant of this pair, respectively.

Lengths and potential N-linked glycosylation sites of the V1–V5* env *sequences. The median length of the V1–V5 *env* sequences from the mothers' viruses was 339 amino acids (IQR, 336–344 amino acids), which was not significantly different from the median length of the same sequences from the infants (337 amino acids; IQR, 336–343 amino acids) ($P = .64$) in the whole group of 17 pairs or according to the timing of transmission (figure 3*B*). Similarly, we observed no difference when individual variables or constant regions were compared separately (data not shown).

There was a median of 22 PNGS in the V1–V5 region (figure 3*C*), which was not significantly different between clones isolated from the mothers (IQR, 21–24 PNGS) and those isolated from the infants (IQR, 22–23 PNGS) ($P = .83$), both overall as well as when the analysis was restricted to cases of in utero transmission ($P = .40$) or to cases of intrapartum transmission ($P = .31$).

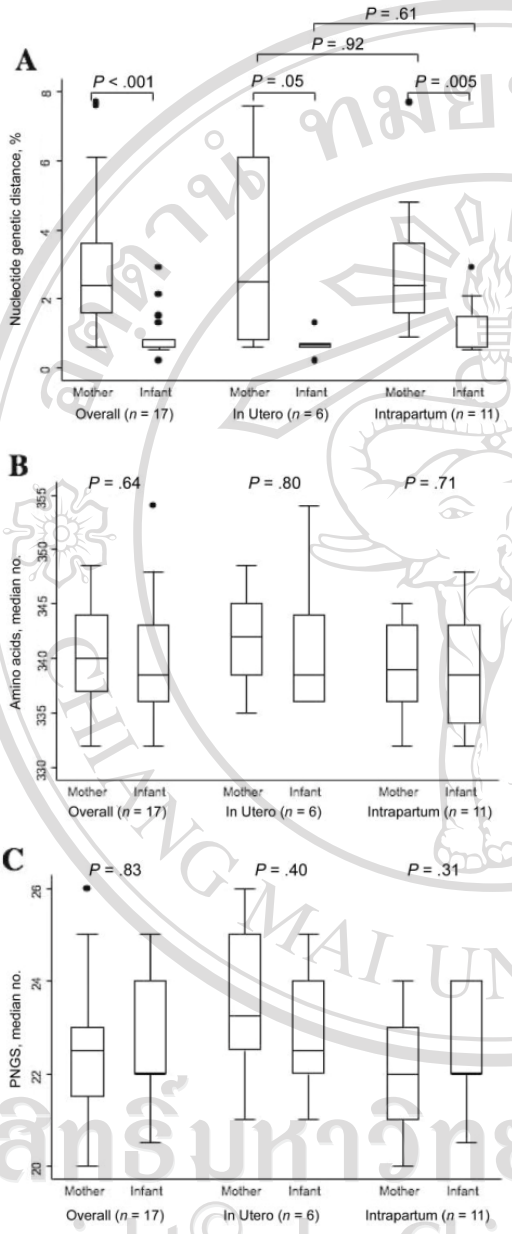


Figure 3. Characteristics of the viral variants in mothers and infants. *A,* Nucleotide intrapatient genetic distances. *B,* Median V1–V5 lengths, expressed as the numbers of amino acids, in mothers and infants. *C,* Median numbers of potential N-glycosylation sites (PNGS) in mothers and infants, in all pairs ($n = 17$), in cases of in utero transmission only ($n = 6$), and in cases of intrapartum transmission only ($n = 11$). The *P* value was calculated using a mixed-effects model (see Methods). The horizontal bar in the box denotes the median value; the lower and upper hinges, the 25th and 75th percentiles; the adjacent lines, the lower adjacent values; and the points, the outside value.

The numbers of PNGS were similar for viruses from mothers and infants, but some positions were highly conserved in all sequences, whereas others were more variable. We therefore tried to map the conserved positions in viruses from the mothers and in viruses from the infants (figure 4). Eight PNGS, including N157 and N161 in the V1–V2 junction; N197, N262, N276, and N289 in C2; N333 in C3; and N440 in C4 were conserved in all viruses from mothers and infants, suggesting an important role in the maintenance of viral integrity. Conversely, a few PNGS were identified as signatures of isolates infecting a given pair, because they were found in all the clones from both mother and infant. These PNGS include, for instance, an additional site at the C-terminus of V1 in pair 779 and N434 in C4 of pair 0357.

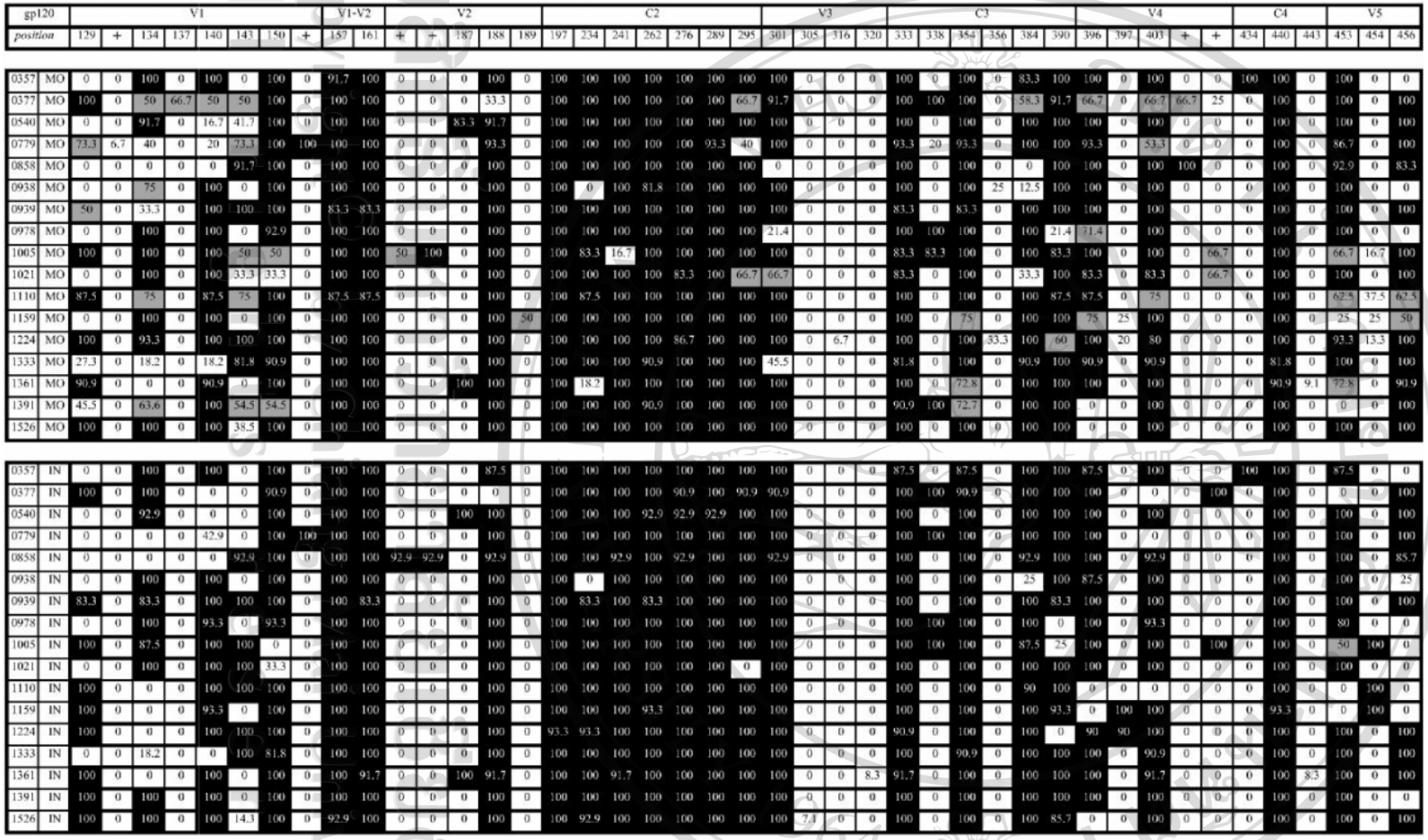
In a few instances, PNGS were found at a high frequency in unusual positions in a few variants in infants only—not in the corresponding virus population in mothers; this finding suggests that these PNGS might confer an advantage to the transmitted virus. Examples of this situation are shown for pairs 1159 and 1224, in which >90% of the infant sequences harbor an additional PNGS at N397 in V4, and for pairs 1005, 1110, and 1159, in which 100% of the infant sequences harbor N454 in V5.

A few PNGS were conserved in almost all infant virus sequences, whereas they were conserved but were present at a lower frequency in the maternal viruses. This was particularly marked for the PNGS at N301. This PNGS was present in 93%–100% of clones from the infants of pairs 0858, 0978, 1021, and 1333, whereas it was totally absent from all 12 clones from the mother in pair 0858 and was present in only 21%, 67%, and 46% of clones from mothers in pairs 0978, 1021, and 1333, respectively. Another PNGS, N384 in C3, displayed a similar pattern, except in pair 0938, in which it was found in only 25% of the infant sequences. The logistic regression model showed that the risk of finding a PNGS was significantly higher ($P < .001$) in the infant than in the mother at the following 4 positions: 301, 397, 454, and 384, as sorted according to decreasing values of the odds ratio of the association. This finding suggests that PNGS N301 and N384, which are shared by many CRF01_AE isolates, might confer a selective advantage on any transmitted virus of this clade.

DISCUSSION

We analyzed 353 sequences encompassing almost the entire gp120 *env* gene from 17 mother-infant pairs infected with HIV-1 CRF01_AE strains, to investigate the characteristics of the envelope genes involved in the MTCT of HIV-1. Several previous studies with similar aims have been performed [12, 13, 15, 16, 22–26, 39], but our study is probably one of the most extensive and homogeneous.

All the sets of sequences from infants clustered closely with the corresponding maternal sequences, but they could be distinguished from the maternal sequences in 14 of the 17 pairs, sug-



gp120		V1								V1-V2		V2				C2								V3				C3						V4						C4			V5		
position		129	+	134	137	140	143	150	+	157	161	+	+	187	188	189	197	234	241	262	276	289	295	301	305	316	320	333	338	354	356	384	390	396	397	403	+	+	434	440	443	453	454	456
0357	MO	0	0	100	0	100	0	100	0	91.7	100	0	0	0	100	0	100	100	100	100	100	100	100	100	0	0	0	100	0	100	0	83.3	100	100	0	100	0	0	100	100	0	100	0	0
0377	MO	100	0	50	66.7	50	50	100	0	100	100	0	0	0	33.3	0	100	100	100	100	100	100	66.7	91.7	0	0	0	100	100	100	0	58.3	91.7	66.7	0	66.7	66.7	25	0	100	0	100	0	100
0540	MO	0	0	91.7	0	16.7	41.7	100	0	100	100	0	0	83.3	91.7	0	100	100	100	100	100	100	100	100	0	0	0	100	0	100	0	100	100	100	0	100	0	0	0	100	0	100	0	100
0779	MO	73.3	6.7	40	0	20	73.3	100	100	100	100	0	0	0	93.3	0	100	100	100	100	100	93.3	40	100	0	0	0	93.3	20	93.3	0	100	100	93.3	0	53.3	0	0	0	100	0	86.7	0	100
0858	MO	0	0	0	0	0	91.7	100	0	100	100	0	0	0	100	0	100	100	100	100	100	100	100	0	0	0	0	100	0	100	0	0	100	100	0	100	100	0	0	100	0	92.9	0	83.3
0938	MO	0	0	75	0	100	0	100	0	100	100	0	0	0	100	0	100	0	100	81.8	100	100	100	100	0	0	0	100	0	100	25	12.5	100	100	0	100	0	0	0	100	0	100	0	0
0939	MO	50	0	33.3	0	100	100	100	0	83.3	83.3	0	0	0	100	0	100	100	100	100	100	100	100	100	0	0	0	83.3	0	83.3	0	100	100	100	0	100	0	0	0	100	0	100	0	100
0978	MO	0	0	100	0	100	0	92.9	0	100	100	0	0	0	100	0	100	100	100	100	100	100	100	21.4	0	0	0	100	100	100	0	100	21.4	71.4	0	100	0	0	0	100	0	100	0	0
1005	MO	100	0	100	0	100	50	50	0	100	100	50	100	0	100	0	100	83.3	16.7	100	100	100	100	100	0	0	0	83.3	83.3	100	0	100	83.3	100	0	100	0	66.7	0	100	0	66.7	16.7	100
1021	MO	0	0	100	0	100	33.3	33.3	0	100	100	0	0	0	100	0	100	100	100	100	83.3	100	66.7	66.7	0	0	0	83.3	0	100	0	33.3	100	83.3	0	83.3	0	66.7	0	100	0	100	0	100
1110	MO	87.5	0	75	0	87.5	75	100	0	87.5	87.5	0	0	0	100	0	100	87.5	100	100	100	100	100	100	0	0	0	100	0	100	0	100	87.5	87.5	0	75	0	0	0	100	0	62.5	37.5	62.5
1159	MO	0	0	100	0	100	0	100	0	100	100	0	0	0	100	50	100	100	100	100	100	100	100	100	0	0	0	100	0	75	0	100	100	75	25	100	0	0	0	100	0	25	25	50
1224	MO	100	0	93.3	0	100	100	100	0	100	100	0	0	0	100	0	100	100	100	100	86.7	100	100	100	0	6.7	0	100	0	100	33.3	100	60	100	20	80	0	0	0	100	0	93.3	13.3	100
1333	MO	27.3	0	18.2	0	18.2	81.8	90.9	0	100	100	0	0	0	100	0	100	100	100	90.9	100	100	100	45.5	0	0	0	81.8	0	100	0	90.9	100	90.9	0	90.9	0	0	0	81.8	0	100	0	100
1361	MO	90.9	0	0	0	90.9	0	100	0	100	100	0	0	100	100	0	100	18.2	100	100	100	100	100	100	0	0	0	100	0	72.8	0	100	100	100	0	100	0	0	0	90.9	9.1	72.8	0	90.9
1391	MO	45.5	0	63.6	0	100	54.5	54.5	0	100	100	0	0	0	100	0	100	100	100	90.9	100	100	100	100	0	0	0	90.9	100	72.7	0	100	100	0	0	100	0	0	0	100	0	0	0	100
1526	MO	100	0	100	0	100	38.5	100	0	100	100	0	0	0	100	0	100	100	100	100	100	100	100	100	0	0	0	100	0	100	0	100	100	0	0	100	0	0	0	100	0	100	0	100
0357	IN	0	0	100	0	100	0	100	0	100	100	0	0	0	87.5	0	100	100	100	100	100	100	100	100	0	0	0	87.5	0	87.5	0	100	100	87.5	0	100	0	0	100	100	0	87.5	0	0
0377	IN	100	0	100	0	0	0	90.9	0	100	100	0	0	0	0	0	100	100	100	100	90.9	100	90.9	90.9	0	0	0	100	100	90.9	0	100	100	100	0	0	0	100	0	100	0	0	0	100
0540	IN	0	0	92.9	0	0	0	100	0	100	100	0	0	100	100	0	100	100	100	92.9	92.9	92.9	100	100	0	0	0	100	0	100	0	100	100	100	0	100	0	0	0	100	0	100	0	100
0779	IN	0	0	0	0	42.9	0	100	100	100	100	0	0	0	100	0	100	100	100	100	100	100	100	100	0	0	0	100	100	100	0	100	100	100	0	0	0	0	0	100	0	100	0	100
0858	IN	0	0	0	0	0	92.9	100	0	100	100	92.9	92.9	0	92.9	0	100	100	92.9	100	92.9	100	100	92.9	0	0	0	100	0	100	0	92.9	100	100	0	92.9	0	0	0	100	0	100	0	85.7
0938	IN	0	0	100	0	100	0	100	0	100	100	0	0	0	100	0	100	0	100	100	100	100	100	100	0	0	0	100	0	100	0	25	100	87.5	0	100	0	0	0	100	0	100	0	25
0939	IN	83.3	0	83.3	0	100	100	100	0	100	83.3	0	0	0	100	0	100	83.3	100	83.3	100	100	100	100	0	0	0	100	0	100	0	100	83.3	100	0	100	0	0	0	100	0	100	0	100
0978	IN	0	0	100	0	93.3	0	93.3	0	100	100	0	0	0	100	0	100	100	100	100	100	100	100	100	0	0	0	100	100	100	0	100	0	100	0	93.3	0	0	0	100	0	80	0	0
1005	IN	100	0	87.5	0	100	100	0	0	100	100	0	0	0	100	0	100	100	100	100	100	100	100	100	0	0	0	100	100	100	0	87.5	25	100	0	100	0	100	0	100	0	50	100	0
1021	IN	0	0	100	0	100	100	33.3	0	100	100	0	0	0	100	0	100	100	100	100	100	100	0	100	0	0	0	100	0	100	0	100	100	100	0	100	0	0	0	100	0	100	0	0
1110	IN	100	0	0	0	100	100	100	0	100	100	0	0	0	100	0	100	100	100	100	100	100	100	100	0	0	0	100	0	100	0	90	100	0	0	0	0	0	0	100	0	0	100	0
1159	IN	100	0	0	0	93.3	0	100	0	100	100	0	0	0	100	0	100	100	100	93.3	100	100	100	100	0	0	0	100	0	100	0	100	93.3	0	100	100	0	0	0	93.3	0	0	100	0
1224	IN	100	0	0	0	100	100	100	0	100	100	0	0	0	100	0	93.3	93.3	100	100	100	100	100	100	0	0	0	90.9	0	100	0	100	0	90	90	100	0	0	0	100	0	100	0	100
1333	IN	0	0	18.2	0	0	100	81.8	0	100	100	0	0	0	100	0	100	100	100	100	100	100	100	100	0	0	0	100	0	90.9	0	100	100	100	0	90.9	0	0	0	100	0	100	0	100
1361	IN	100	0	0	0	100	0	100	0	100	91.7	0	0	100	91.7	0	100	100	91.7	100	100	100	100	100	0	0	8.3	91.7	0	100	0	100	100	100	0	91.7	0	0	0	100	8.3	100	0	100
1391	IN	100	0	100	0	100	0	100	0	100	100	0	0	0	100	0	100	100	100	100	100	100	100	100	0	0	0	100	0	100	0	100	100	100	0	100	0	0	0	100	0	100	0	0
1526	IN	100	0	100	0	100	14.3	100	0	92.9	100	0	0	0	100	0	100	92.9	100	100	100	100	100	100	7.1	0	0	100	0	100	0	100	85.7	0	0	100	0	0	0	100	0	100	0	100

Figure 4. Map of potential N-glycosylation sites (PNGS) for each quasispecies present in either mother or infant. We recorded the positions of PNGS in the V1–V5 sequences from each of the 353 clones, with respect to the CRF01_AE consensus sequence. We then calculated the frequency of each PNGS in the sequences of each mother or infant. PNGS were considered to be conserved at a given position if they were present in >80% of the clones from a given individual. The positions of the PNGS correspond to those of the consensus HIV-1 CRF01_AE obtained from the HIV databases of the Los Alamos National Laboratory [35]. Additional PNGS found in our subjects are denoted by a plus sign (+). The frequency of PNGS at each position is presented as a percentage for each case. Mothers are denoted by an ID number followed by "MO," and infants are denoted by an ID number followed by "IN." Values >80% are highlighted in black, values of 50%–79.9% are highlighted in gray, and those <50% are highlighted in white.

gesting a significant bottleneck at transmission. In these 14 infants, viral sequences were highly homogeneous, with a genetic distance much shorter than that of the viruses in the corresponding mothers. Multiple variants were clearly shown to be transmitted from mother to infant in only 1 pair, with similar large genetic distances noted between the variants from the infant and the variants from the mothers. Phylogenetic analysis of the *env* sequences from the mothers and infants of the 2 remaining pairs showed these sequences to be intermingled. However, there was a limited genetic distance between the maternal sequences, suggesting that evolution of the virus was restricted in these mothers. It is therefore unclear whether these 2 cases can really be classified as multiple variant transmission. We found no difference in genetic diversity between viruses present in infants infected in utero and viruses present in infants infected during the intrapartum period; however, because of the small sample size, the power to detect such differences was only 25% in mothers and 16% in infants. This result is consistent with the findings of some [16, 24, 26] but not all [24] previous studies.

Extensive diversity was observed in the mother *env* sequences, whereas the majority of infant samples showed limited viral diversity. We believe that this difference was not the cause of resampling bias during the PCR steps, particularly because there was a high plasma VL in most of the infants and because we used the entire amount of extracted RNA for reverse transcription. Therefore, this suggests that there was no limitation in the number of viral templates during the first round of PCR. In addition, we performed multiple, subsequent, independent PCRs for several of these infant samples, which were pooled before cloning and sequencing. A similar limited viral diversity was observed. Finally, additional studies performed on the contemporary infant PBMCs showed that the same limited viral diversity was found in both PBMC samples and plasma samples (data not shown).

Studies of viruses isolated in early infection are of prime importance, because these viruses correspond to variants capable of successfully establishing infection in natural conditions. Recent studies have highlighted potential differences in the biology of sexual transmission between viral subtypes [27–29, 40]. Viruses of subtypes A and C appear to pass through a genetic bottleneck during or shortly after heterosexual transmission, leading to the selection of viruses with a compact and less glycosylated gp120 [27, 28]. This molecular property was not observed for sexually transmitted subtype B viruses [28, 29].

Descriptions of the molecular properties of vertically transmitted viruses in infants provide additional information, because these viruses are successfully transmitted to infants despite the presence of preexisting passively transferred antibodies, particularly if transmission occurs during the intrapartum period or late in pregnancy [23]. They therefore provide excellent examples of neutralization escape variants that are potentially able to improve our understanding, for the benefit of future vaccine approaches. On the basis of recent observations suggesting that sexually transmitted viruses have shorter gp120 proteins and fewer PNGS, we examined these sequence characteristics in the gp120 genes of our mother-infant pairs. We observed no difference between the viruses of the mothers and infants, in terms of both the length of the V1–V5 region and the number of PNGS. These findings differ slightly from those of the recent study by Wu et al. [23], in which the V1–V5 lengths of the envelope did not differ between infants and mothers but in which fewer PNGS were found in viruses from infants than in those from their mothers, who were infected with viruses of various subtypes. Although our study focused on a non–breast-feeding group, whereas the study of Wu et al. [23] focused on a breast-feeding population, this discrepancy may also be the result of differences in the biology of different subtypes, and it highlights the need to perform such studies in various continents, with the use of viruses from diverse clades.

The numbers of PNGS were similar in the viruses from mothers and infants in our study, but we nonetheless analyzed their location in gp120. The dense carbohydrate (glycan) array surrounding the HIV envelope protein is now known to evolve continually during natural infection—a powerful adaptation for escape from neutralizing antibodies and for virus survival in the host [41–43]. We therefore mapped the PNGS that were conserved or variable in all mother-infant pairs. The most striking observation was the high degree of conservation of 2 PNGS in viruses in infants, at positions N301 in V3 and N384 in C3. These PNGS were significantly less conserved in maternal viruses ($P < .001$, by a logistic regression model). This pattern suggests that these PNGS may be associated with a selective advantage, at least for CRF01_AE viruses. Interestingly enough, the N301 glycan site was also found in all maternally transmitted CRF01_AE viruses sequenced in an independent study in Thailand [44]. In addition, several in vitro studies have shown that N301 is associated with both a decrease in the sensitivity of HIV-1 to neutralization by CD4BS antibodies and modulation of the interaction of the HIV-1 envelope with CD4 and chemokine receptors [45–47].

In conclusion, we analyzed the molecular characteristics of transmitted HIV-1 viruses of the CRF01_AE clade in MTCT. We also provided evidence that, despite the presence of a complex virus population in the mother, only a restricted subset of viruses is transmitted to the infant, regardless of whether transmission occurs in utero or in the intrapartum period. We did not find that more-compact (i.e., shorter) gp120 regions or fewer PNGS were characteristic of viruses transmitted from mother to infant, as has been suggested for at least a few subtypes of sexually transmitted viruses. However, our data suggest that a limited number of PNGS that appear to be conserved in all infant variants may confer an advantage on the virus to be transmitted, probably allowing it to escape antibody-mediated neutralization, as suggested by the most recent studies in the field of MTCT [22, 23].

Functional studies are required to confirm these findings, but our results nonetheless highlight the role of the "glycan shield" in HIV biology [48]. Dissection of the molecular characteristics of viruses transmitted to infants in MTCT, in which exposure to the virus occurs in the presence of passively transferred preexisting antibodies, could provide key elements for the prevention of HIV infection through vaccination or passive immunization.

Acknowledgment

We thank René Ecochard, Laboratoire Biostatistique, Santé Université Claude Bernard, Lyon I (Unité Mixte de Recherche Centre National de la Recherche Scientifique 5558), Lyon, France, for his advice regarding the statistical analysis.

References

1. John GC, Kreiss J. Mother-to-child transmission of human immunodeficiency virus type 1. Epidemiol Rev **1996**; 18:149–57.
2. Kourtis AP, Lee FK, Abrams EJ, Jameson DJ, Bulterys M. Mother-to-child transmission of HIV-1: timing and implications for prevention. Lancet Infect Dis **2006**; 6:726–32.
3. Thorne C, Newell ML. Mother-to-child transmission of HIV infection and its prevention. Curr HIV Res **2003**; 1:447–62.
4. Lallemant M, Jourdain G, LeCoeur S, et al. Single-dose perinatal nevirapine plus standard zidovudine to prevent mother-to-child transmission of HIV-1 in Thailand. N Engl J Med **2004**; 351:217–28.
5. McIntyre J. Strategies to prevent mother-to-child transmission of HIV. Curr Opin Infect Dis **2006**; 19:33–8.
6. Ho DD, Neumann AU, Perelson AS, Chen W, Leonard JM, Markowitz M. Rapid turnover of plasma virions and CD4 lymphocytes in HIV-1 infection. Nature **1995**; 373:123–6.
7. Preston BD, Dougherty JP. Mechanisms of retroviral mutation. Trends Microbiol **1996**; 4:16–21.
8. Rambaud A, Posada D, Crandall KA, Holmes EC. The causes and consequences of HIV evolution. Nat Rev Genet **2004**; 5:52–61.
9. Frost SDW, Wrin T, Smith DM, et al. Neutralizing antibody responses drive the evolution of human immunodeficiency virus type 1 envelope during recent HIV infection. Proc Natl Acad Sci USA **2005**; 102:18514–9.
10. Shankarappa R, Margolick JB, Gange SJ, et al. Consistent viral evolutionary changes associated with the progression of human immunodeficiency virus type 1 infection. J Virol **1999**; 73:10489–502.
11. Burton DR, Stanfield RL, Wilson IA. Antibody vs. HIV in a clash of evolutionary titans. Proc Natl Acad Sci USA **2005**; 102:14943–8.
12. Wolinsky SM, Wike CM, Korber BT, et al. Selective transmission of human immunodeficiency virus type-1 variants from mothers to infants. Science **1992**; 255:1134–7.
13. Scarlatti GT, Leitner E, Halapi J, et al. Comparison of variable region 3 sequences of human immunodeficiency virus type 1 from infected children with the RNA and DNA sequences of the virus populations of their mothers. Proc Natl Acad Sci USA **1993**; 90:1721–5.
14. Ahmad N, Baroudy BM, Baker RC, Chappey C. Genetic analysis of human immunodeficiency virus type 1 envelope V3 region isolates from mothers and infants after perinatal transmission. J Virol **1995**; 69:1001–12.
15. Pasquier C, Cayrou C, Blancher A, et al. Molecular evidence for mother-to-child transmission of multiple variants by analysis of RNA and DNA sequences of human immunodeficiency virus type 1. J Virol **1998**; 72: 8493–501.
16. Verhofstede C, Demecheleer E, De Cabooter N, et al. Diversity of the human immunodeficiency virus type 1 (HIV-1) *env* sequence after vertical transmission in mother-child pairs infected with HIV-1 subtype A. J Virol **2003**; 77:3050–7.
17. Scarlatti G, Albert J, Rossi P, et al. Mother-to-child transmission of human immunodeficiency virus type 1: correlation with neutralizing antibodies against primary isolates. J Infect Dis **1993**; 168:207–10.
18. Lathey JL, Tsou J, Brinker K, Hsia K, Meyer WA III, Spector SA. Lack of autologous neutralizing antibody to human immunodeficiency virus type 1 (HIV-1) and macrophage tropism are associated with mother-to-infant transmission. J Infect Dis **1999**; 180:344–50.
19. Bongertz V, Costa CI, Veloso VG, et al. Neutralization titres and vertical HIV-1 transmission. Scand J Immunol **2002**; 56:642–4.
20. Guevara H, Casseb J, Zijenah LS, et al. Maternal HIV-1 antibody and vertical transmission in subtype C virus infection. J Acquir Immune Defic Syndr **2002**; 29:435–40.
21. Barin F, Jourdain G, Brunet S, et al. Revisiting the role of neutralizing antibodies in mother-to-child transmission of the HIV-1. J infect Dis **2006**; 193:1504–11.
22. Dickover R, Garratty E, Yusim K, Miller C, Korber B, Bryson Y. Role of maternal autologous neutralizing antibody in selective perinatal transmission of human immunodeficiency virus type 1 escape variants. J Virol **2006**; 80:6525–33.
23. Wu X, Parast AB, Richardson BA, et al. Neutralization escape variants of human immunodeficiency virus type 1 are transmitted from mother to infant. J Virol **2006**; 80:835–44.
24. Dickover RE, Garratty EM, Plaeger S, Bryson YJ. Perinatal transmission of major, minor, and multiple maternal human immunodeficiency virus type 1 variants in utero and intrapartum. J Virol **2001**; 75:2194–203.
25. Lamers SL, Sleasman JW, She JX, et al. Persistence of multiple maternal genotypes of human immunodeficiency virus type I in infants infected by vertical transmission. J Clin Invest **1994**; 93:380–90.
26. Renjifo B, Chung M, Gilbert P, et al. *In-utero* transmission of quasispecies among human immunodeficiency virus type 1 genotypes. Virology **2003**; 307:278–82.
27. Derdeyn CA, Decker JM, Bibollet-Ruche F, et al. Envelope-constrained neutralization-sensitive HIV-1 after heterosexual transmission. Science **2004**; 303:2019–22.
28. Chohan B, Lang D, Sagar M, et al. Selection for human immunodeficiency virus type 1 envelope glycosylation variants with shorter V1-V2 loop sequences occurs during transmission of certain genetic subtypes and may impact viral RNA levels. J Virol **2005**; 79:6528–31.
29. Frost SDW, Liu Y, Pond SL, et al. Characterization of human immunodeficiency virus type 1 (HIV-1) envelope variation and neutralizing antibody responses during transmission of HIV-1 subtype B. J Virol **2005**; 79:6523–7.
30. Li B, Decker JM, Johnson RW, et al. Evidence for potent autologous neutralizing antibody titers and compact envelopes in early infection with subtype C human immunodeficiency virus type 1. J Virol **2006**; 80: 5211–8.
31. Lallemant M, Jourdain G, Le Coeur S, et al. A trial of shortened zidovudine regimens to prevent mother-to-child transmission of human immunodeficiency virus type 1. Perinatal HIV Prevention Trial (Thailand) Investigators. N Engl J Med **2000**; 343:982–91.
32. Bryson YJK, Luzuriaga JL, Sullivan JL, Wara DW. Proposed definitions for *in utero* versus intrapartum transmission of HIV-1. N Engl J Med **1992**; 327:1246–7.
33. Hall TA. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucl Acids Symp Ser **1999**; 41:95–8.
34. Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position specific gap penalties and weight matrix choice. Nucleic Acids Res **1994**; 22:4673–80.
35. Los Alamos National Laboratory. HIV databases. Available at: http://www.hiv.lanl.gov/content/index. Accessed 13 July 2008.
36. Kumar S, Tamura K, Nei M. MEGA3: integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment. Brief Bioinform **2004**; 5:150–63.

37. Los Alamos National Laboratory. N-GlycoSite. Available at: http://www.hiv.lanl.gov/content/sequence/GLYCOSITE/glycosite.html. Accessed 13 July 2008.
38. Kimura M. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J Mol Evol **1980**; 16:111–20.
39. Tovanabutra S, de Souza M, Sittisombut N, et al. HIV-1 genetic diversity and compartmentalization in mother/infant pairs infected with CRF01_AE. AIDS **2007**; 21:1050–3.
40. Blish CA, Nedellec R, Mandaliya K, Mosier DE, Overbaugh J. HIV-1 subtype A envelope variants from early in infection have variable sensitivity to neutralization and to inhibitors of viral entry. AIDS **2007**; 21: 693–702.
41. Dacheux L, Moreau A, Ataman-Onal Y, Biron F, Verrier B, Barin F. Evolutionary dynamics of the glycan shield of the human immunodeficiency virus envelope during natural infection and implication for exposure of the 2G12 epitope. J Virol **2004**; 78:12625–37.
42. Scanlan CN, Offer J, Zitzmann N, Dwek RA. Exploiting the defensive sugars of HIV-1 for drug and vaccine design. Nature **2007**; 446:1038–45.
43. Wei X, Decker JM, Wang S, et al. Antibody neutralization and escape by HIV-1. Nature **2003**; 422:307–12.
44. Sutthent R, Foongladda S, Chearskul S, et al. V3 sequence diversity of HIV-1 subtype E in infected mothers and their infants. J Acquir Immune Defic Syndr Hum Retrovirol **1998**; 18:323–31.
45. Kiszka I, Kmieciak D, Gzyl J, et al. Effect of the V3 loop deletion of envelope glycoprotein on cellular responses and protection against challenge with recombinant vaccinia virus expressing gp160 of primary human immunodeficiency virus type 1 isolates. J Virol **2002**; 76:4222–32.
46. Koch M, Pancera M, Kwong PD, et al. Structure-based, targeted deglycosylation of HIV-1 gp120 and effects on neutralization sensitivity and antibody recognition. Virology **2003**; 313:387–400.
47. Malenbaum SE, Yang D, Cavacini L, Posner M, Robinson J, Cheng-Mayer C. The N-terminal V3 loop glycan modulates the interaction of clade A and B human immunodeficiency virus type 1 envelopes with CD4 and chemokine receptors. J Virol **2000**; 74:11008–16.
48. Van Kooyk Y, Engering A, Lekkerkerker AM, Ludwig IS, Geijtenbeek TBH. Pathogens use carbohydrates to escape immunity induced by dendritic cells. Curr Opin Immunol **2004**; 16:488–93.

III. Intrapatient Recombination between Human Immunodeficiency Virus Type 1 Envelope Genes in Mother-to-Child Transmission

1. Background and specific aims

We conducted the previous study to describe the molecular characteristics of the HIV-1 variants involved in MTCT, based on sequencing of a large number of envelope gene sequences from seventeen mother-child pairs [the study part 2; 353]. We found that the length of gp120 and number of potential N-linked glycosylation sites (PNGS) were not characteristics of viruses transmitted from mother to infant, unlike suggested for at least a few subtypes in the situation of sexual transmission [79; 107; 131 ; 236]. However, our data suggested that a limited number of PNGS that appear to be conserved in all infant variants might confer an advantage on the virus to be transmitted [the study part 2; 353]. A further in depth analysis of all the sequences allowed us to identify two cases of transmission of intrapatient recombinants from the mother to their infant. In this part of our study, we report these two cases that suggest that recombination probably contributed to adaptation of HIV-1 to its environment to be successfully transmitted from mothers to their infants. This is to our knowledge the first description of the contributing role of HIV recombination in MTCT. In addition, our data allow both to confirm, in natural *in vivo* conditions, a hot spot for recombination in the C2 region of *env*, and to suggest another hot spot in the C3 region.

Specific aim was to identify the intrapatient recombination between HIV-1 envelope genes that may contribute to successfully transmit HIV-1 from mothers to their infants.

2. Materials and methods

Patient material

We previously analyzed 353 envelope clones covering almost the entire gp120 region of the *env* gene (V1-to-V5) from 17 mother-child pairs infected with HIV-1 CRF01_AE clade in Thailand [this study, part 2; 353]. Infants were not breastfed. In that previous study, proviral DNA of mothers taken at the time of delivery and plasma virus-derived cDNA of infants taken at, or as close as possible to the first time point at which HIV-1 DNA PCR was positive, were used as templates for PCR amplification of the HIV-1 *env* gene. Amino acid alignments of *env* clones and comparison of intra-pair sequences pointed 2 pairs (# 1005 and 0779) among 17 in

which either V1-V2 or V3 regions of *env* genes from the infant's viruses were similar to two distinct phylogenetic groups of maternal variants. This was suggestive of potential intrapatient recombination. In order to identify those recombinations, additional samples from both mothers and infants were taken at different time points, depending of the availability of biological material (Table 5). In summary, peripheral blood mononuclear cells (PBMCs) and plasma were available at delivery for each mother. PBMCs were available at birth and plasma was available at day 47 for infant 1005. Based on positive detection of HIV DNA at birth, infant 1005 was considered as infected *in utero* [56]. HIV DNA detection was negative at birth but positive at day 124 for infant 0779, indicative of intra-partum infection [56]. Plasma and PBMCs from infant 0779 were available at day 124 and 224, respectively.

Table 5. Samples used during the present study.

ID pair	Mother or Infant	Date of delivery/ birth (d/m/y)	Date of first HIV DNA positive by PCR (d/m/y)	Date of sampling (d/m/y)	Time point (days after birth)	Source of sequences	Nb of clones	HIV-1 RNA (log copies/ mL)
1005	Mother	28/01/00		28/01/00	0	DNA	22	4.16
				28/01/00	0	RNA	9	
	Infant		28/01/00	28/01/00	0	DNA	10	3.87
				15/03/00	47	RNA	17	ND*
0779	Mother	4/06/99		4/06/99	0	DNA	21	3.45
				4/06/99	0	RNA	9	
	Infant		6/10/99	6/10/99	124	RNA	11	5.01
				14/01/00	224	DNA	10	ND*

*Not determined

d/m/y: day/month/year

Amplification, cloning and sequencing.

DNA was extracted from PBMCs using the QIAamp DNA Blood Midi kit (Qiagen). HIV-1 RNA was extracted from 280 µl plasma using the QIAamp Viral RNA Mini Kit (Qiagen) and cDNA was synthesized with the SuperScript First-Strand Synthesis System for RT-PCR (Invitrogen), using random hexamers. A 1.2 kb fragment of upstream V1 to downstream V5 was amplified by nested-PCR (Platinum PCR SuperMix High Fidelity; Invitrogen) with subtype-specific primers AEs1ext, AEa1ext, AEs1int, and AEa1int [this study, part 2; 353]. The PCR products were cloned using the pCR2.1-TOPO TA cloning kit (Invitrogen) and sequenced with the Big Dye Terminator V3.1 Cycle Sequencing kit (Applied Biosystems) and an ABI 3100 genetic analyzer (Applied Biosystems), using six overlapping primers encompassing the entire gp120 sequence as described [this study, part 2; 353]. PCR

amplifications were performed separately for each individual sample, to avoid cross contaminations. For both maternal proviral DNA and infant viral RNA of each pair, clones that were obtained and sequenced in the previous study were included in the present analysis. However, a full series of extraction, amplification, and cloning was performed again for the same samples in order to get an additional set of clones for each of them. This allowed us to confirm the original findings in two sets or independent experiments.

Phylogenetic analysis.

Nucleotide sequences were assembled and translated into amino acid sequences with the BioEdit package [158], aligned with Clustal W [394] and manually adjusted to optimize the alignment. Sequences with premature stop codon were excluded. A total of 109 clones were analyzed. Neighbor-joining trees were constructed with MEGA version 4 [387] with bootstrap analysis using 1,000 replicates.

Prediction of coreceptor usage.

The co-receptor usage of all clones was predicted based on V3 sequences using various approaches: the 11/25 rule which is taking into account the key residues 11 and 25 and total net charge of V3 loop [105; 126], Geno2pheno [coreceptor] which is the bioinformatics tool based on a support vector machines (SVM) [374], WebPSSM which is a tool based on a position-specific scoring matrix (PSSM) [175], and R5/X4-pred which is a tool based on a random forest (RF) [435]. In case of discrepancy, we privileged the conclusion of the RF method because it was developed considering non-B non-C viruses [143; 435].

Recombination analysis.

To identify recombinants, our data sets were split into three regions (V1-V2, V3, and V4-V5) and the phylogenetic trees were constructed for each variable region for each mother-child pair [269; 351]. A clone was considered a recombinant if it clustered with different phylogenetic groups of sequences separated by significant bootstrap values (≥ 90%) at least in two trees. Putative parental sequences were identified as the sequences most similar to the recombinant in these trees. Then, we performed a BootScan analysis to confirm these findings and to identify the recombination breakpoints (Simplot; [240]). In order to facilitate the analysis, the sequences derived from the PBMCs DNA were grouped according to the phylogenetic analysis for each mother. Maternal proviral sequences were considered as belonging to one group if they clustered together systematically in the three trees obtained with the three variable regions by significant bootstrap values. There were four and five groups of maternal proviral DNA sequences for pairs 1005 and 0779, respectively. A consensus sequence, in which each amino-acid is that most frequently present at a given position, was defined for sequences from the infants, sequences

issued from the maternal plasma viral RNA, and for each of the groups of maternal PBMCs sequences. Simplot software was then used to compare each clone from the infant sequences or from the maternal viral RNA (the query sequence) with the maternal consensus sequences of each group (the reference sequences). The analyses were also done using the maternal consensus sequence of each group as the query and the consensus sequences of the other maternal groups as the references. This allowed us to identify the genetic organization of each viral population within each pair. In addition, the recombinant and parental sequences were inspected manually.

Nucleotide sequence accession numbers.

All sequences were deposited in GenBank under accession numbers EU031392-EU031405, EU031258-EU031279, and EU784739-EU784811.

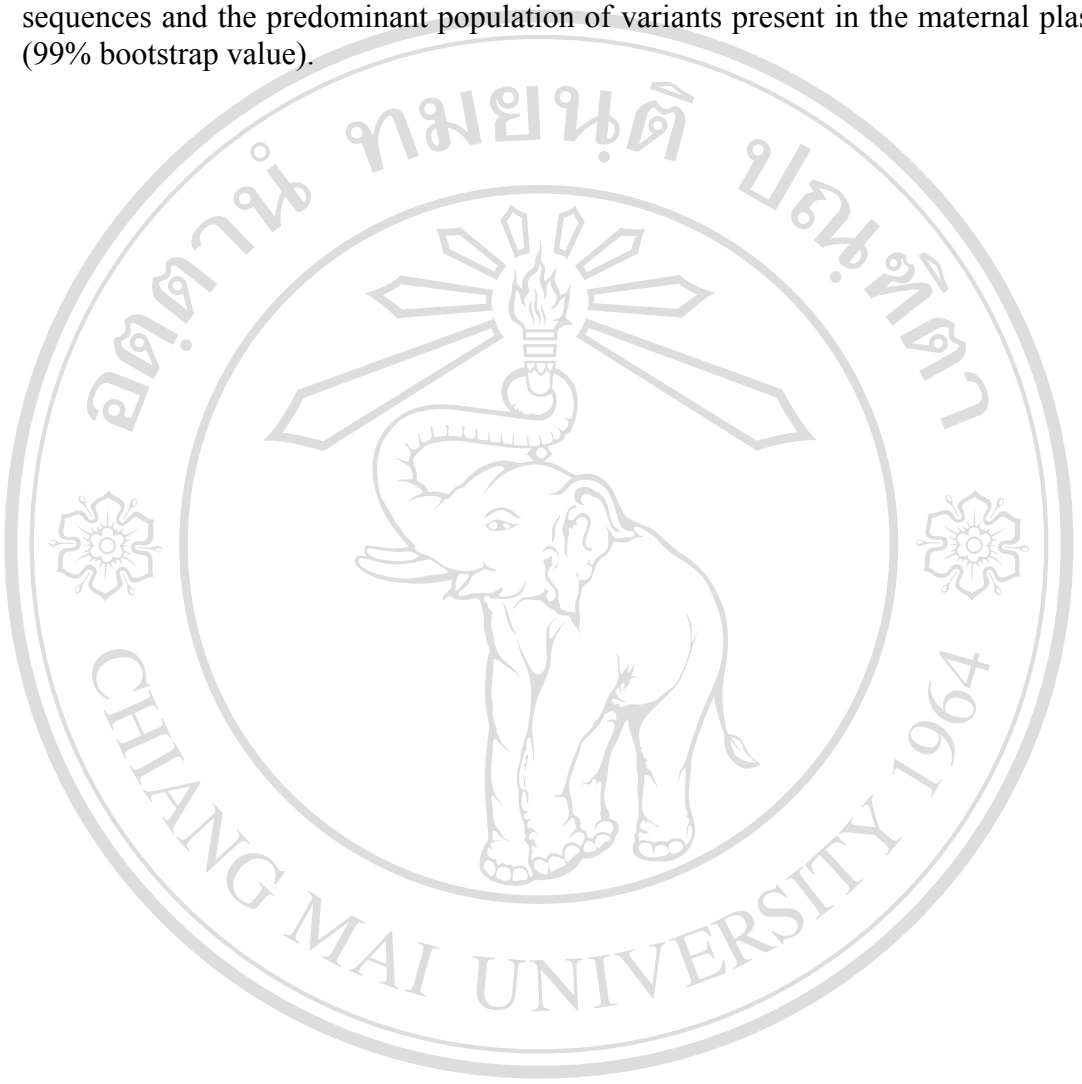
3. Results

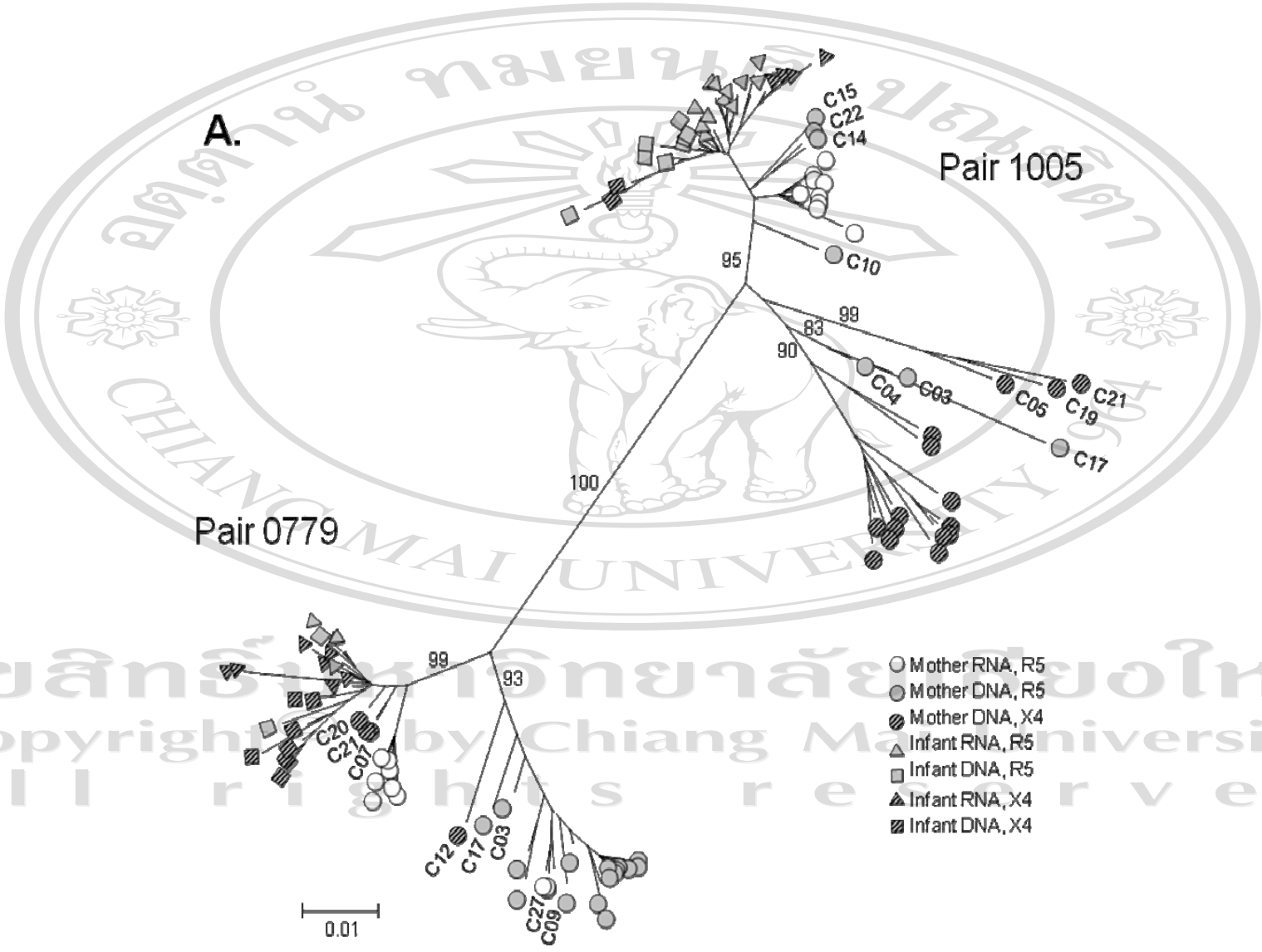
Phylogenetic analysis and prediction of co-receptor usage of viral populations in two mother-infant pairs.

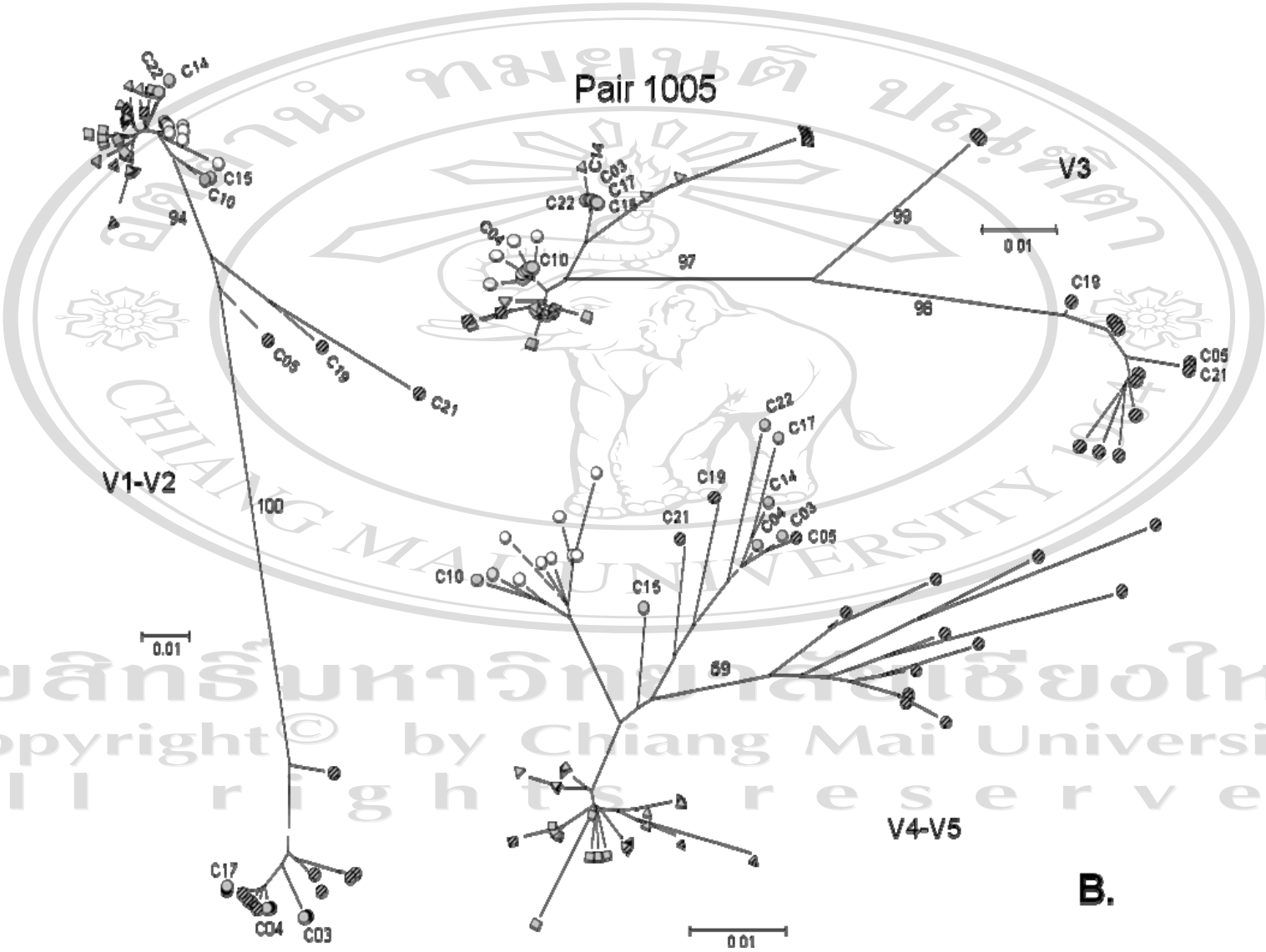
To investigate the relationship between the maternal and infant sequences, entire gp120 (V1-V5) sequences of 58 and 51 envelope clones of pair 1005 and 0779, respectively, were analyzed. Figure 40A presents the phylogenetic tree for the two mother-child pairs. The sequences from each mother-infant pair clustered together with 100% bootstrap value, confirming the epidemiological link between each mother and her child, and indicating the absence of cross-patient contamination. In both pairs, the infant's viral sequences were closely related to those present in the maternal plasma at delivery. Only a few clones issued from the maternal PBMCs DNA clustered with the maternal RNA sequences in each case (clones C07, C20, and C21 for pair 0779; clones C10, C14, C15, and C22 for pair 1005), whereas most of clones derived from the maternal PBMCs DNA were highly diverse, probably representative of archived older variants. The plasma viral RNA sequences of mother 0779 were separated in two populations (7 and 2 clones, respectively) suggesting two groups of viral variants circulating at time of delivery.

The co-receptor usage of all clones was predicted based on V3 sequences using various approaches. The four prediction methods were in total agreement for more than 90% of the clones. The results are shown in Figure 40 A-C. In pair 1005, most of the infant sequences (21 of 27 clones) and all 9 maternal sequences issued from the plasma RNA were predicted as R5, whereas the majority of maternal DNA sequences (15 of 22 clones) were predicted as X4. Four clones issued from the maternal PBMCs DNA (C10, C14, C15, C22) that clustered with the infant sequences and the maternal viral RNA sequences (95% bootstrap value), were also predicted as R5. The opposite was observed in pair 0779 in which the majority of infant sequences (15 of 21 clones) were predicted as X4, whereas the majority of maternal DNA sequences (17 of 21

clones) and all 9 maternal RNA sequences were predicted as R5. Three of 4 maternal DNA clones (C07, C20 and C21) that were predicted as X4 clustered with the infant sequences and the predominant population of variants present in the maternal plasma (99% bootstrap value).







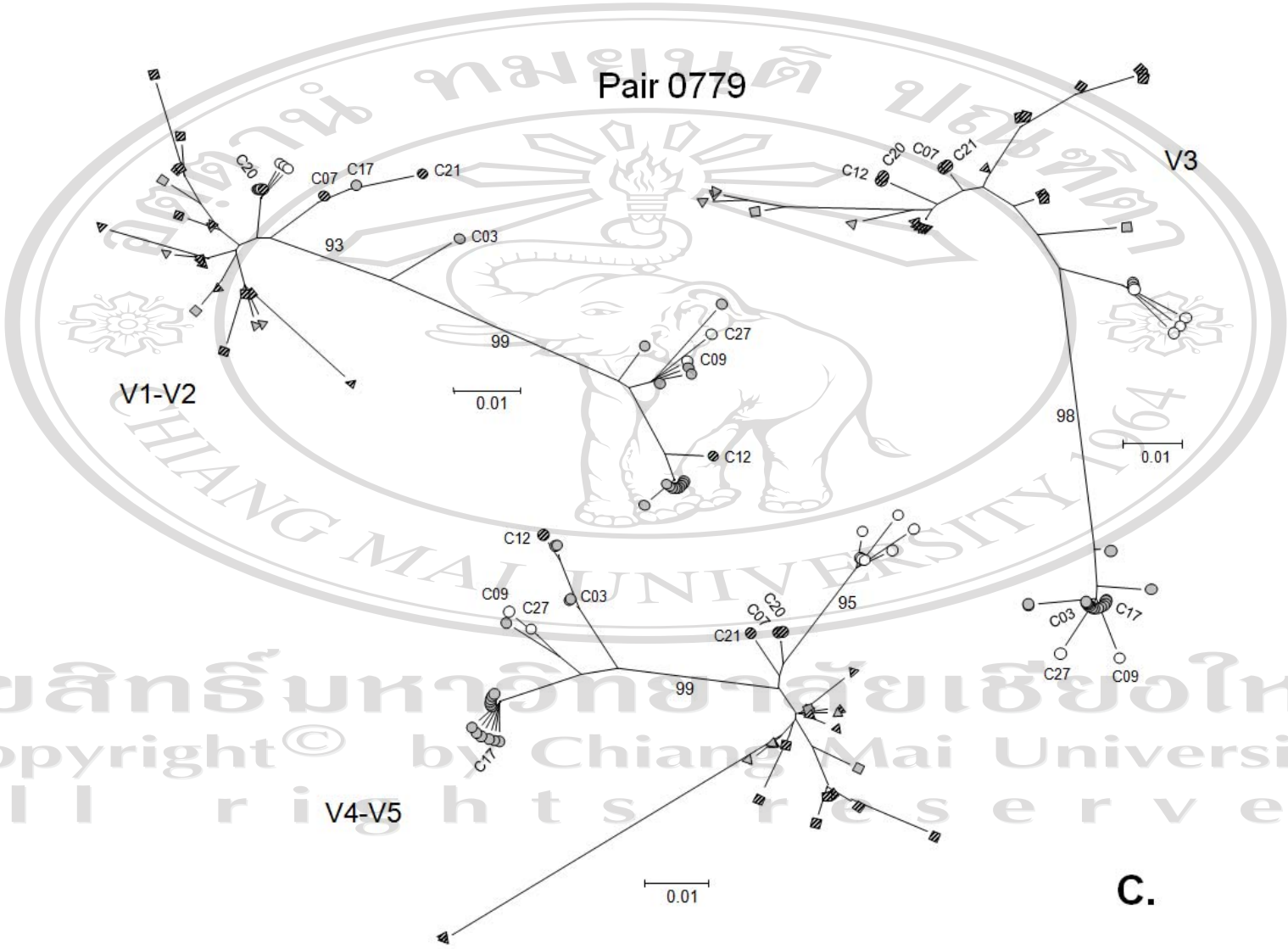


Figure 40. Unrooted neighbor-joining trees of the 109 HIV-1 *env* gp120 nucleotide sequences from the two mother-child pairs. Phylogenetic trees were constructed from:

(**A**) the V1-V5 region (almost entire gp120 sequence) from pairs 1005 and 0779;

(**B**) the V1-V2, V3, and V4-V5 regions from pair 1005;

(**C**) the V1-V2, V3, and V4-V5 regions from pair 0779.

Horizontal branch lengths are drawn to scale, with the bar representing 1% divergence. Bootstrap values are indicated as percentages per 1,000 replicates on branches. Each symbol represents a single *env* sequences. Open circles indicate plasma RNA-derived maternal sequences, grey circles represent PBMCs DNA-derived maternal sequences, squares represent PBMCs DNA-derived infant sequences, and triangles represent plasma RNA-derived infant sequence. Hatched symbols show clones predicted as X4 (see legend more precisely in A). The numbered clones are those cited in the text.

Evidence for intra-mother recombination and transmission to the infant.

To identify recombinants, our data sets were split into three regions (V1-V2, V3, and V4-V5) and the phylogenetic trees were constructed for each variable region for each mother-child pair (Figure 40, B and C). Since recombinants have acquired genetic material from at least two sources, they should cluster with different groups of sequences when trees are constructed from subsets of the data [269; 351]. Therefore, a clone was considered a recombinant if it clustered with different phylogenetic groups of sequences separated by significant bootstrap values (≥ 90%) at least in two trees. This is exemplified in pair 1005 where the maternal DNA clones 03, 04, and 17, clustered with both the infant clones and the maternal clones issued from the plasma viral RNA in the V3 tree, but not in the V1-V2 tree nor the V4-V5 tree (Figure 40B). Similarly, clone 12 of mother 0779 clustered with the infant clones in the V3 region whereas it was much more distantly related in the V1-V2 and V4-V5 trees (Figure 40C). The infant sequences and all maternal RNA sequences (except two clones, C09 and C27, for mother 0779) clustered in the same branch with high bootstrap values in all trees for both pairs, providing additional proof of similarity between infant viruses and the predominant variant present in the plasma of their mothers. Taken together, these data suggested that the infant sequences and the maternal plasma sequences were the results of recombination events between envelope genes that were archived in the maternal PBMCs DNA.

We next performed a BootScan analysis to confirm these findings and to identify the recombination breakpoints. Simplot software was used to compare each clone from the infant sequences or from the maternal viral RNA (the query sequence) with the maternal consensus sequences of each group (the reference sequences). The analyses were also done using the maternal consensus sequence of each group as the query and the consensus sequences of the other maternal groups as the references. The results are presented in figure 41 (A and B) where both the BootScan analysis using the infant viral RNA consensus sequence as the query, and the summary of the genomic organization of each viral population are shown. The BootScan results indicated the crossover points at the same positions at nucleotides 278, 463, and 750 for all clones of pair 1005, and at nucleotides 252, 347, and 738 for all clones of pair 0779. Therefore two breakpoints were located in the C2 region and one breakpoint was located in the C3 region. Interestingly, the recombination breakpoint in C3 region was at the exactly same position for both pairs (nucleotide position 718 – starting at T of the first codon TGC of V1 - of the CRF01_AE consensus sequence from the HIV database; http://hiv.lanl.gov) (Figure 42). In pair 1005, the mosaic virus patterns of infant sequences, maternal RNA sequences and maternal DNA sequences group 4, were similar whereas the patterns of each maternal DNA sequences (groups 1 to 4) were different (Figures 41A and 42A). Within the maternal DNA sequences, the V1-V2 region was similar between groups 1 and 2, and groups 3 and 4; the V3 region was identical between groups 1 and 3, and groups 2 and 4; and the V4-V5 was similar between groups 1 and 4, and groups 2 and 3 (Figure 42A). This indicates that the recombination occurred within virus populations that have been archived in the PBMCs, and that the maternal DNA sequences related to group 4

might be representative of the contemporaneous maternal circulating population that was transmitted to the infant. In pair 0779, the mosaic patterns of infant sequences, maternal RNA sequences and maternal DNA sequences group 5, were similar whereas the patterns of each group of maternal DNA sequences (groups 1 to 5) were different (Figures 41B and 42B). The evidence for recombination was found within maternal DNA sequences group 1, 3, 4, and 5. The V1-V2 region was identical between groups 1 and 3, and groups 4 and 5; the V3 region was identical between groups 1 and 4, and groups 3 and 5. The group 2 sequences had a more complex composition, the V3 region being similar to groups 1 and 4 (Figure 42B). The maternal DNA sequences of group 5 were closely related to both the infant viral sequences and the predominant maternal circulating sequences; however their predicted co-receptor usage was different (Figure 41B). Although this was only prediction and therefore should be confirmed by phenotypic assays, this different tropism might be due to two mutations, R/S and V/I at positions 11 and 30 of the V3 region, respectively. R11 and V30 were present in both the group 5 population of the mother and the infant virus, but were replaced by S11 and I30 in the maternal plasma viral population. It has been demonstrated clearly that substitutions at position 11 are sufficient to switch coreceptor usage [104; 127].

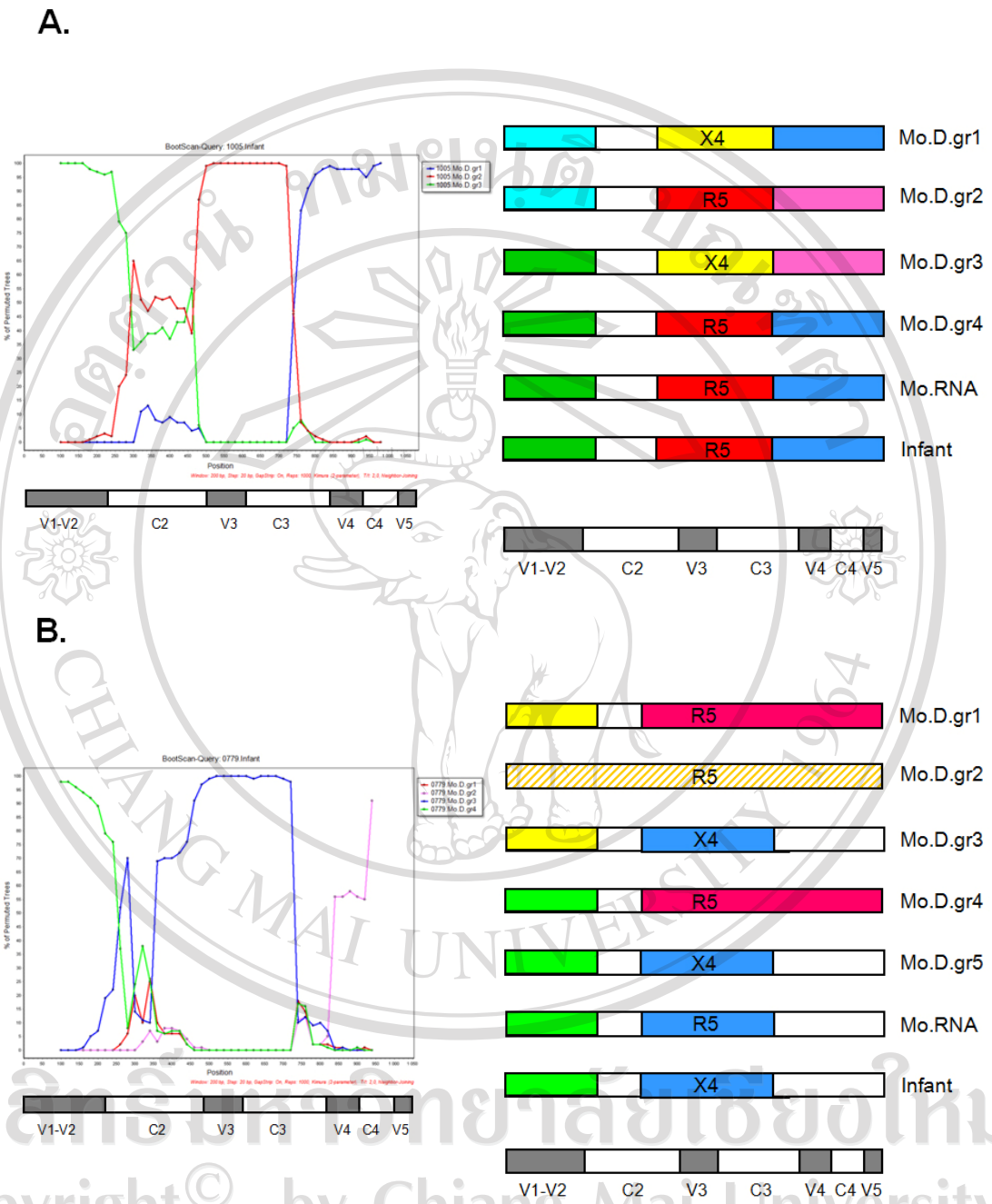


Figure 41. Schematic representation of the recombination analysis for pairs 1005 (**A**) and 0779 (**B**). Left panel: representative BootScan analysis using the sequence of one clone from the infants' viral RNA as query and the maternal consensus sequences of each group as references. Right panel: schematic representation of the recombinant patterns derived from the BootScan analysis for each group of sequences. The constant and variable regions of HIV-1 *env* gp120 are shown at the bottom. The colors presented in the left panel correspond to those in the right panel. The predicted coreceptor usage (R5 or X4) is indicated for each group of sequences. Mo (mother); D (DNA).

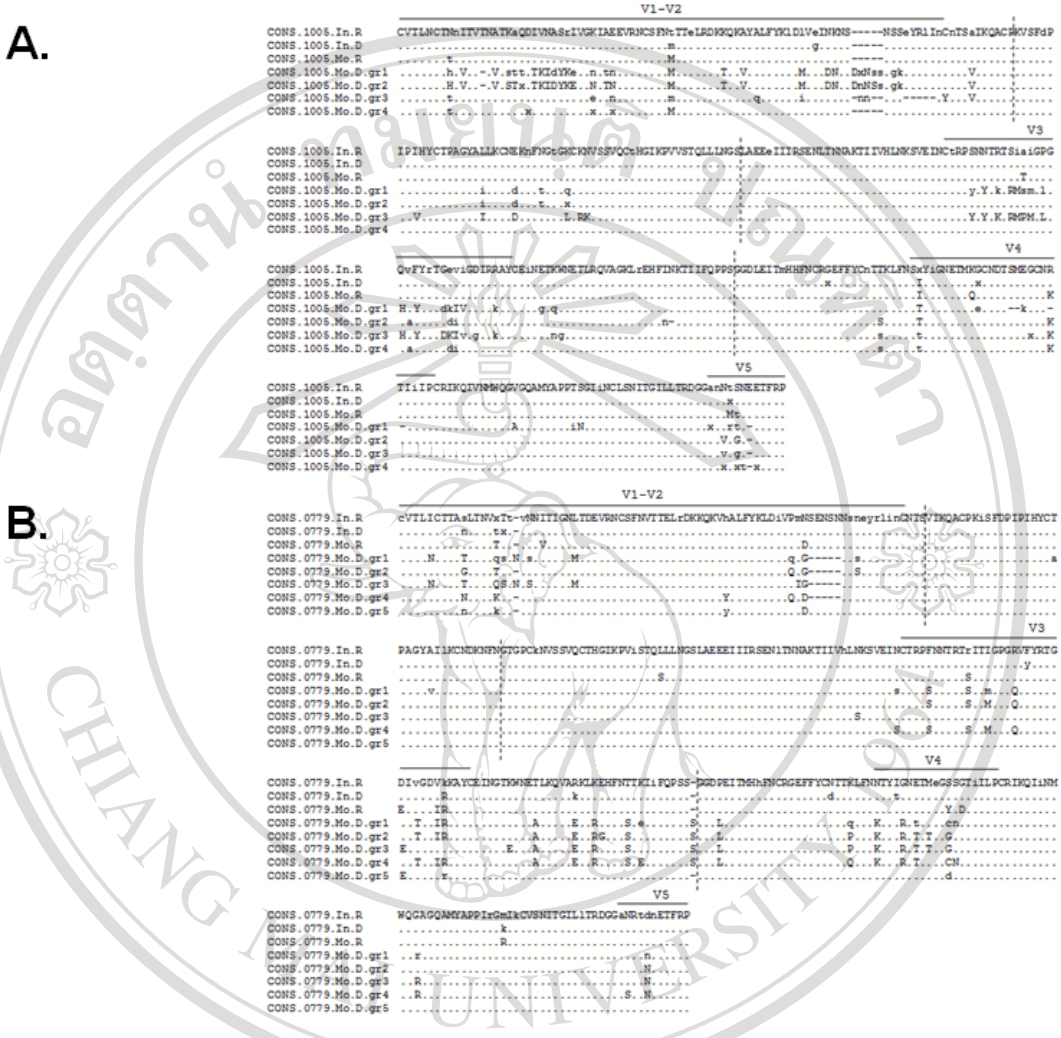


Figure 42. Alignment of HIV-1 *env* gp120 (V1-V5) amino acid consensus sequences of each group for pairs 1005 (**A**) and 0779 (**B**). Uppercase letters indicate amino acids conserved in 100% of sequences of each group. Lowercase letters indicate amino acids conserved in more than 50% of sequences of each group. 'x' is any amino acid. Dots represent identical amino acids. Dashes represent deletions. In.R. corresponds to the consensus sequence of clones derived from the infants viral RNA. In.D. corresponds to the consensus sequence of clones derived from the infants PBMCs DNA. Mo.R. corresponds to the consensus sequence of clones derived from the maternal viral RNA. Mo.D. corresponds to the consensus sequence of clones derived from the maternal PBMCs DNA (for each group). The vertical dot lines show the location of the recombinant breakpoints.

4. Discussion

The study of a large number of HIV-1 *env* gene sequences from 17 mother-child pairs infected with CRF01_AE variants [this study, part 2; 353] led us to identify two cases of intrapatient recombinations. Recombinants may be generated *in vitro* during the process of PCR [324]. Several arguments make *in vitro* recombination an unlikely explanation for the origin of the recombinants reported here. First, the same recombinant profile was obtained in both mother and infant for each pair. Second, the same recombinant was found in both the plasma viral RNA and the PBMCs proviral DNA for each case, even when plasma and PBMCs were collected at different time periods as it was the case for the infants. Third, the clones obtained from the amplified proviral DNA from the mothers and those obtained from the amplified plasma viral RNA from the infants were derived from two fully independent sets of experiments performed at several months interval, including different extractions, amplifications and clonings. The same recombinants were recovered in clones derived from the two independent series. Taken together, these observations support that the recombinant sequences originated *in vivo*.

To our knowledge, this is the first report showing that intrapatient recombinations within the HIV-1 *env* gp120 gene that occur in mothers are involved in transmission of HIV-1 to their children. This would suggest that, in these two cases, the recombinant variants might have acquired a selective advantage that made them adapted to MTCT. One could imagine that recombination might be implicated in generation of viruses able to evade the passively transferred maternal neutralizing antibodies, and/or able to interact more efficiently with the receptor or coreceptors at the surface of target cells. Interestingly enough, the recombinant viruses transmitted to the infants for both pairs harbored V1-V2 and V3 regions issued from different parental sequences archived in the proviral DNA. This suggests that not only V1-V2 region nor V3 region alone, but both of them act additionally or synergistically to provide a specific advantage in selection of the transmitted variant. This was previously demonstrated for the coreceptor switching [69; 154; 253; 269; 302], and it has been shown that the V1-V2 region and V3 region are important for inducing neutralizing antibody responses [148; 280; 350; 429; 431; 438; 444]. The implication of recombination events in MTCT was also suggested recently in a study performed in Tanzanian women. That study indicated that viruses with intersubtype recombinant LTRs were more likely to be transmitted through breastfeeding than viruses with non-recombinant form of subtype A, C or D [208].

In both cases of our report, the transmitted virus was homogeneous, genetically restricted, and closely related to the viral population that was prevalent in the maternal plasma at delivery. The comparison of all the sequences and the results from the BootScan analysis showed that the recombinations occurred between different phylogenetic groups of sequences that were archived in the maternal DNA. Interestingly, the recombinant breakpoints were located in both pairs in the C2 and C3 regions. The recombination in C2 is in agreement with previous data that identified the C2 region as a hotspot for recombination [29; 135]. Surprisingly, the recombinant

breakpoint in the C3 region was identical in both pairs, suggesting that this position is also a hotspot for recombination, at least in the CRF01_AE clade (Figure 42).

The two cases reported herein provide naturally occurring biological material that may be very useful to document how the association of variable regions initially present on different genomes, when associated following a recombination event, may lead to an advantage in selective transmission of MTCT of HIV-1.

5. Publications

These data have been submitted to the Journal of Virology in August, 2008.

Samleerat T, Moreau A, Ngo-Giang-Huong N, Jourdain G, Leechanachai P, Pornkitprasarn W, Bhakeecheep S, Lallemant M, Barin F. *Intrapatient Recombination between Human Immunodeficiency Virus Type 1 Envelope Genes in Mother-to-Child Transmission.*

This work has been accepted for presentation at The AIDS Vaccine 2008 Conference. Cape Town, South Africa. October 13-16, 2008.

Samleerat T, Moreau A, Ngo-Giang-Huong N, Jourdain G, Leechanachai P, Pornkitprasarn W, Bhakeecheep S, Lallemant M, Barin F. *Intrapatient recombination in human immunodeficiency virus type 1 envelope involved in mother-to-child transmission (MTCT) of HIV-1 subtype CRF01_AE*. The AIDS Vaccine 2008 Conference. Cape Town, South Africa. October 13-16, 2008 ***(poster presentation-presenting author)***.