Characterization of envelope sequence of HIV virus in children infected with HIV in Vietnam

Linh Vu Phuong Dang[1] [iD]**, Hung Viet Pham[2], Thanh Thi Dinh[1], Thu Hoai Nguyen[3], Quyen Thi Huyen Vu[3], Nhung Thi Phuong Vu[1], Phuong Thi Bich Le[1], Lam Van Nguyen[3], Hai Thanh Le[3], Phuong Thi Vu[4]* and Linus Olson[5,6,7]***

Abstract

Background: HIV is characterized by high levels of genetic variability, including increased numbers of heterogeneous sequences of the envelope region. Therefore, studying genetic variability of HIV in relation to viral replication might facilitate prognosis of disease progression.

Methods: The study was designed as cross-sectional; data and samples of participants collected and analyzed *env* genes were obtained from 23 children enrolled by Vietnam National Children's Hospital.

Results: Substantial mutations in the C2 region were found in patients with high levels of viral replication while changes in the C3 region were mostly found in patients with low viral load. In the V1 region, we found profound amino acid modifications in patients with low HIV viral loads in contrast to the V2 sequence, where we identified single point mutations in patients with increased HIV viral load. The V3 region was relatively homogeneous, while profound deletions in the V4 region were detected in patients with increased viral replication.

Conclusion: Our results suggest that genetic variations in different regions of the HIV envelope sequence, including both conserved C2 and C3 and variable V1/V2 and V4 regions, might be involved in increased viral infectivity and replication capacity. Such knowledge might help improve prediction of HIV progress and treatment in patients.

Keywords

HIV, envelope, CD4 T cell counts, HIV viral load, children, Vietnam

Introduction

According to the UNAIDS report, 2.1 million children have been infected with HIV (human immunodeficiency virus) worldwide, of which more than 6500 HIV-infected children were found in Vietnam.[1,2] The majority of the infected children become infected with HIV vertically, through mother-to-child transmission (MTCT). Thus, HIV might already encounter higher levels of genetic variability. The perinatal HIV infection is influenced by a combination of factors including both maternal, infant host, and viral factors. While the genetic polymorphisms in human leukocyte antigen (HLA) genes and chemokine receptor genes in mother and infant host have been shown to play a crucial role in MTCT of HIV, the genetic variability of HIV virus in relation to

[1]Hanoi University of Public Health, Hanoi, Vietnam
[2]Vietnam National Children's Hospital, Hanoi, Vietnam
[3]Ninh Binh Maternity Hospital, Hanoi, Vietnam
[4]Dinh Tien Hoang Institute of Medicine, Hanoi, Vietnam
[5]Training and Research Academic Collaboration (TRAC) – Sweden – Vietnam
[6]Department of Global Public Health, Karolinska Institutet, Stockholm, Sweden
[7]Department of Woman's and Children's Health, Karolinska Institutet, Stockholm, Sweden

*Shared last authors.

Corresponding author:
Linh Vu Phuong Dang, Hanoi University of Public Health, 1A Duc Thang Road, North Tu Liem District, Hanoi, 1000, Vietnam.
Email: phuonglinh.j@gmail.com or dvpl@huph.edu.vn

MTCT has not been studied extensively.[3,4] In addition, it has been shown that the genetic variation due to heterosexual transmission could be different compared to other transmissions, including men who have sex with men (MSM) transmissions and MTCT. Therefore, genetic variability could be crucial factors in order to determine not only the successful transmission but also the course of the HIV infection.[5]

Genetic variability of HIV occurs as the result of nucleotide substitutions, deletions, or insertions within each genotype or recombination between different genotypes. Four major genotype groups of HIV have been reported, including M (main), O (outlier), N (non-M, non-O), and P (putative) groups.[6] Among all the groups, group M, the most predominant circulating genotype, has been divided into different subtypes (clades) denoted A, B, C, D, F, G, H, J, and K and sub-subtypes denoted A1, A2, A3, A4, F1, F2, and so on.[7] The variation within subtype could be 15%–20%, while up to 25%–30% difference has been reported[8] between subtype genetic variation. In addition to the single subtypes, researchers have reported other recombinant subtypes, which are classified as circulating recombinant forms (CRFs) and unique recombinant forms (URFs),[9] thought to be the results of combination between subtypes in dually infected persons. These subtypes could pass to other people through different routes of transmission.[10] Subtype AE, which also is defined as CRF01_AE, has been shown to be the most frequent in Vietnam in different groups, regardless of the routes of transmission.[11,12]

HIV infects CD4 T cells through binding of its envelope (env) protein gp120 to the CD4 receptor and the CXCR4/CCR5 co-receptor on the cell surface. After binding, HIV releases its genome into the host cells, in which the HIV genome undergoes reverse transcription into DNA, which is then inserted into the host's genome in the infected cells. The inserted DNA is then transcribed and translated to form new HIV virions including the whole HIV genome and associated proteins. HIV virions then use the host cells' surface to form the viruses which are later released and able to infect other cells.[13–15] The process has multiple steps and involves complex pathways, yet it is believed that the fitness of HIV is critical for viral replication, transmission, and disease progression. Hence, HIV has been shown to be completely dependent upon the env protein to enter the host cells.

The env gene encodes the surface-expressed viral protein env (gp160) and linearly contains five conserved regions (C1–C5) and five variable regions (V1–V5).[16] Among the C regions, the C2 and C3 regions have been shown to be crucial for CD4 binding. C3 contains three residues defined as oligosaccharide sites that are required for structure and function of the protein. These residues consist of N-X-S (Asparagine-X-Serine) or N-X-T (Asparagine-X-Threonine) contexts including NGT (Asparagine-Glycine-Threonine), NVS (Asparagine-Valine-Serine), and NGS (Asparagine-Glycine-Serine).[17] Regarding the V region, the loop structure was shown to be formed by the intermolecular disulfide bonds in the gp120 glycoprotein.[18] The V1 region plays a

significant role in early infection.[19,20] Both V1 and V2 regions are associated with increased neutralization susceptibility of HIV.[21,22] V3 is one of the most important determinants of viral tropism and co-receptor usage with the involvement of N-X-S (Asparagine-X-Serine) or N-X-T (Asparagine-X-Threonine) contexts.[23] The functional importance of V3 is illustrated by the fact that deletion of V3 completely abrogates virus infectivity.[24] The role of V4 is unclear regarding HIV infection; it has been shown to be involved in formation of neutralizing antibodies.[25]

It has been suggested that HIV has high levels of genetic variability, and thus that evolution of the *env* gene might occur rapidly, during the course of HIV infection.[26,27] The process, in turn, could facilitate viral replication by increased fitness, and increased ability to escape from the selection pressure caused by anti-retroviral therapy (ART), as well as the immune system. However, other mutations in the *env* gene might be associated with decreased viral entry and disturbance of certain functions of host cells. Quan et al.[28] have reported that mutations in the *env* gene, particularly G367R and D368R in the CD4 binding site (CD4bs) of gp120, have led to inactivity of viral infection.

To our knowledge, studies of diversifications of the *env* gene in the CRF01_AE clade of HIV, the predominant clade circulating in Vietnam, are limited. Here we investigate variations of the env region in HIV-infected children registered at Vietnam National Children's Hospital, Hanoi, Vietnam, and its relation to HIV viral load and CD4 T cell counts.

Methods

Study population and sample collection

The study subjects were HIV-infected children, who have been diagnosed and treated at National Hospital of Pediatrics (NHP), Hanoi, Vietnam, in 2012. Samples chosen from a previous study[29] were those taken during the first year after treatment initiation and they were compared to the variability of the *env* gene among HIV-infected children who were treated with first-line ART. Among samples we have accessed, we could extract RNA and perform further analysis in 23 samples. The data regarding CD4 T cell counts and HIV viral load were collected from medical records of all patients at the same time point as the one when blood samples were taken. The duration of the study was from August 2017 to October 2017.

Polymerase chain reaction, cloning, and sequencing of HIV-1 env

Blood samples were collected in tubes containing ethylenediaminetetraacetic acid (EDTA) and transported to the laboratory within 6 h to perform RNA extraction. Viral RNA extraction was performed using QIAamp Viral RNA Mini Kit (Qiagen, Hilden, Germany) and purified RNA was used to synthesize cDNA using First-Strand Synthesis System for

Table 1. General characteristics of HIV-infected children.

Patients	Age	Sex	CD4 T cell counts	HIV RNA	Opportunistic infection	Clinical stage
1014	4.7	M	680	400	Y	1
1022	5.1	F	754	400	Y	1
1027	2.8	F	407	1848	Y	3
1032	1.8	M	29	220	Y	3
1033	3.0	F	888	400	N	1
1035	5.0	M	236	220	Y	4
1037	4.7	M	1754	400	Y	1
1058	1.6	M	470	480,000	Y	1
1060	5.1	F	356	400	Y	2
1061	4.0	M	1106	400	Y	3
1065	4.0	F	663	2600	Y	3
1071	4.3	M	114	400	Y	2
1072	7.8	M	884	42,000	Y	2
1081	5.7	M	272	160,000	Y	2
1082	0.9	F	32	180,000	Y	2
1085	9.5	F	374	16,000	Y	2
1087	4.2	F	54	1,300,000	Y	3
1098	5.5	M	594	400	Y	2
1106	6.2	M	740	56,000	Y	2
1125	4.6	F	55	208,193	Y	2
1165	9.2	M	543	400	Y	2
1175	2.2	M	585	86,000	Y	2
1185	3.0	F	400	400	Y	2

Characteristics	Median	Interquartile range (IQR)	
Age	4.6	3	5.5
Weight	15	11	16.5
Height	98	90	103
CD4 T cell counts	470	103	740
HIV RNA	400	400	86,000

reverse transcriptase (RT)–polymerase chain reaction (PCR) (Invitrogen, Carlsbad, CA, USA). The *env* gene was amplified by PCR using a nested-PCR protocol in a PerkinElmer 480 Thermal Cycler (Perkin Elmer, Waltham, MA, USA). PCR amplification of the complete gp160-encoding sequence (the region from 5580 to 8586 of the HIV-1HXB2 genome) was performed using the following external primers: AEs1ext (5′-TGGGTTACAGTTTATTATGGGG-3′) and AEa1ext (5′-TTCTCTTTGCCYTGGTGGGTGCTA -3′) and inner primers including AEs1int (5′-TGCCAAAGC ATATGAGACAGARGYGCA-3′) andAEa1int (5′-TTCAC TTCTCCAATTGTCCTTYATRTT-3′), which have been described previously.[30]

The amplification was carried out in a final volume of $50\,\mu L$ containing Tris-HCl, $MgCl_2$, deoxyribonucleotide triphosphates (dNTPs), primers, and Taq polymerase. Thermal cycle conditions included heating at 94°C for 2 min, at 94°C for 30 s, at 50°C for 30 s, and at 68°C for 90 s for 35 cycles, and a final extension step at 72°C for 10 min for both cycling conditions. The PCR products were detected by agarose gel electrophoresis, followed by ethidium bromide staining. The sequencing

was performed using BigDye terminator reaction kits and an Applied Biosystems 3100 capillary sequencer (Applied Biosystems, Inc., Foster City, CA, USA). The sequences obtained were aligned to a reference sequence, CRF-AE that has been found in Vietnam, using BioEdit software (BioEdit, Tom Hall, Ibis Therapeutics, Carlsbad, CA, USA), and then to blast using the National Center for Biotechnology Information (NCBI) program and the multiple alignment program for amino acid or nucleotide sequences (MAFFT), program version 7 (CBRC). There were certain regions with multiple mutations within the same patients, making it difficult to obtain the clean sequences and hence these were excluded from the analysis.

Results

General characteristics of the HIV-infected children

As shown in Table 1, median age of participants was 4.6 years (interquartile range (IQR): 3–5.5), median CD4 T cell counts were 470 (103–740) cells/mm³ and median

```
ACX31816.1        NITENFNMWKNKMVEQMQEDVISLWDQSLKPCVKLTPLCV
1098              NVTENFNMWKNKMVEQMQEDVISLWDQSLKPCVKLTPLCV
1014              NVTESFNMWKNNMVEQMQEDVISLWDQSLKPCVKLTPLCV
1058              NVTENFNMWKNNMVEQMQEDVISLWDQSLKPCVKLTPLCV
1106              NVTENFNMWKNNMVEQMQEDVISLWDQSLKPCVKLTPLCV
1087              NVTENFNMWKNNMVEQMQEDVISLWDQSLKPCVKLTPLCV
1060              NVTENFNMWKNNMVEQMQEDVISLWDQSLKPCVKLTPLCV
1037              NVTENFDMWKNNMVEQMQEDVISLWDQSLKPCVKLTPLCV
1165              NVTENFNMWKNNMVEQMQEDIISL*DQSLQPCVKLTPLCV
1185              NVTDEFNMWKNNMVEQIHEVVISLLDQSLQLSVQLTTLSV
```

C2 sequence

```
ACX31816.1        GYAILKCNDKNFNGTGPCKNVSSVQCTHGIKPVVSTQLLLNGSLAE
1032              GYAILKCNDKNFNGTGPCKNVSSVQCTHGIKPVVSTQLLLNGSLAE
1037              GYAILKCNDKNFNGTGPCKNVSSVQCTHGIKPVVSTQLLLNGSLAE
1060              GYAILKCNDKNFNGTGPCKNVSSVQCTHGIKPVVSTQLLLNGSLAE
1071              GYAILKCNDKNFNGTGPCKNVSSVQCTHGIKPVVSTQLLLNGSLAE
1087              GYAILKCNDKNFNGTGPCKNVSSVQCTHGIKPVVSTQLLLNGSLAE
1060              GYAILKCNDKNFNGTGPCKNVSSVQCTHGIKPVVSTQLLLNGSLAE
1106              GYAILKCNDKNFNGTGPCRNVSSVQCTHGIKPVVSTQLLLNGSLAE
1027              GYAIIKCNDKNFNGTGPCKNVSSVQCTHGIKPVVSTQLLLNGSLAE
1033              GYAILKCNDKNFNGTGPCKNVSSVQCTHGIKPTVSTQLLLNGSLAE
1082              GYVILKCNDKNFNGTGPCNNVSSVQCTHGIKPVVSTQLLLNGSLAE
1072              GYAILKCNDKNFNGTGPCNNVSSVQCTHGIKPVVSTQLLLNGSLAE
1081              GYVLLKCNDKNFNGTGPCKNVSSVQCTHGIKPVVSTQLLLNGSLAE
1035              GFAILKCNDKRFNGTGPCKNVSSVQCTHGIKPVVSTQLLLNGSLAE
1061              GYALLKCNDKNFNRTGPCKNVSSVQCTHGIKPVVSTQLLLNGSLAE
1014              GYVILKCNDKNFNGKGTCKNVSSVQCTHGIKPVVSTQLLLNGSLAE
1185              GYRMLKCNDKTFNGTGVCSNVSSVQCTHGV-PVVLTKLMLDGGVVV
1022              GYAILKCNDKNFNGTGPCKNVSSVQCTHGHASGINSIAVEWQSRRR
```

C3 sequence

Figure 1. Multi-alignment of C2 and C3 regions of env sequence.

levels of HIV viral load were 400 (400–86,000) copies/ mL. The participants were 43.5% female and 56.5% male. 95.7% of the participants were infected with at least one opportunistic infection and the majority was at clinical stage 2 (52.2%). 30.4% of the HIV-infected children suffered from severe immune-deficiency (CD4 T cell counts < 350 cells/mm^3), 47.8% had high levels of HIV viral load (>1000 copies/mL).

Constant regions showed high homology when aligned to reference sequence

Among 23 subjects included in our study, there were no significant variations in C2 regions compared to the reference sequence. We observed that most of the subjects carried Valine (V) instead of Isoleucine (I) at the second amino acid. Subject 1014, who had low levels of both CD T cell counts and HIV viral load, obtained the replacement of Asparagine (N) to Serine (S) at the fifth amino acid. While high CD4 T cell counts and

low HIV RNA have been found in subject 1037 with alteration from Asparagine (N) to Aspartate (D) at the seventh amino acid, subject 1185 with substantial mutations in 10 amino acids, and subject 1165 with five variations in amino acids (Figure 1).

Comparing with the reference sequence in C3 region, patients (1085 and 1022) with the highest levels of genetic variations had high CD4 T cell counts and low levels of HIV viral load. On the contrary, subjects 1081 and 1085 with high levels of viral load experienced only two variations at positions 3 and 4. While the majority of patients showed high levels of homology in three residues in C3 domains that are defined as oligosaccharide sites required for the structure/function of the protein consisting of N-X-S (Asparagine-X-Serine) or N-X-T (Asparagine-X-Threonine) contexts including NGT (Asparagine-Glycine-Threonine), NVS (Asparagine-Valine-Serine), and NGS (Asparagine-Glycine-Serine), patient 1022 with high CD4 and low HIV RNA showed mutation in these residues (Figure 1).

The low variation in V1 region and single point mutation were found in patient with high levels of viral load

Sequence comparison with the founder virus showed profound amino acid variation in the V1 region in subjects with low viral load in contrast to single point mutations within the V2 region in subjects with increased viral load (Figure 2). Similarly, patients with the full V1 length experienced high levels of HIV viral load in contrast to missing amino acids in the V2 region in participants with high levels of viral replication (Figure 2).

The V3 region remained relatively homogeneous compared to the reference sequence

As can be seen from Figure 2, Cysteine (C), which is responsible for formation of the loop structure, was highly conserved in all patients. In nine patients, four amino acid GPGQ contexts (Glycine-Proline-Glycine-Glutamine) were found to be conserved. The third amino acid was conserved in all patients. While other subjects showed different patterns of variation and viral load at the first and second position, only patients with replacement of Q by R (Arginine) in the last position showed high levels of viral load (Figure 2).

Four patients (1081, 1072, 1175, and 1125) with highly unconserved residues in N-X-S (Asparagine-X-Serine) or N-X-T (Asparagine-X-Threonine) contexts were found to have high levels of HIV viral load in contrast to patients with low levels of diversification who had low HIV RNA (Figure 2).

The most frequent variations found in the V4 region

All subjects showed at least nine mutations in the V4 region, in which subjects 1024, 1071, and 1120 showed the most variations, including multiple point mutations, insertions, and deletions. Five patients had V4 regions with multiple deletions: 1120, 1061, 1130, 1082, and 1125 (Figure 2).

Discussion

Alignment of the C2 region taken from participants to a reference sequence revealed different types of genetic variation. Patients with certain types of genetic variation also experienced high levels of viral load, suggesting that these diversifications might facilitate viral replication in HIV-infected patients.

Regarding the C3 region, patients who carried Serine instead of Asparagine at the fifth amino acid had low CD4 T cell counts and low viral load. The variation might be associated with decrease of HIV fitness; therefore, even though the immune system has not been normalized, HIV replication remained low. On the contrary, patients with substantial

Figure 2. Multi-alignment of V1–V4 regions of env sequence.

variations were found to have high CD4 T cell counts and low levels of HIV viral load, suggesting that the mutation could both decreased fitness and increased immune recognition.

In contrast to the C2 region, significant diversifications in the C3 region were found in patients with increased viral replication, suggesting different roles of the C3 and C2 regions in facilitating viral replication. Higher levels of evolutionary rate in the C2 and C3 regions have been reported in HIV-1 compared to those of HIV-2, and it is suggested that changes in the C2 and C3 regions might be associated with immune escaping mechanisms and viral replication.[17] Moreover, it has been proposed that immune pressure is one of the most important driving forces for rapid intra-host evolution of the C2 and C3 regions. In our study, the evolution of C2 seems to favor viral replication, whereas evolution of C3 appeared to have an opposite role.

The V1 region showed abundant diversifications in HIV-infected children, with truncated regions found in most of the patients. Surprisingly, patients with a full length V1 region showed high levels of HIV viral load, whereas those with a truncated V1 region showed the opposite result. Thus, full sequence of V1 might be required for viral replication. The V1 region has been shown to involve in early infection and replication efficiency[19,20] and deletion of the V1 loop enhanced the neutralization capacity of the immune system.[20] In line with these findings, we showed that the presence of a full length V1 region was associated with increased viral replication and that a truncated V1 region was found in patients with low levels of viral load. However, the mechanism behind the truncation of the V1 region is still unclear, probably due to the increased surveillance by the immune system as long as it is intact.

For the V2 region, among the highly conserved four amino acid GPGQ context, only the variation in the last residue was found in patients with high levels of HIV viral load, implying a role of the last residue in viral replication. The V1/V2 regions are considered to play an important role in escaping the immune response, since deletions in the V1/V2 regions were associated with poor neutralization susceptibility of HIV.[21,22] However, most of the participants with truncated V1 regions in our study had low viral replication, suggesting that the V1 region might have different roles compared to the V2 region in development of neutralization capacity of the hosts. Consistent with our results, Saunder et al.[23] also emphasized different roles of V1 and V2 loops in association with the neutralization susceptibility profile of HIV-1 and in certain cases, the roles of V1 and V2 might even be opposite. Our study appears to favor opposite roles of V1 and V2 for increased viral replication. In addition, V1 and V2 changes have been found to be associated with alteration of co-receptor usage.[31] However, such changes were not consistently found, and have not been documented in the AE subtype. Thus, it is difficult to compare the changes in association with the changes in co-receptor usage and HIV

binding. Trinh et al. showed that alterations of the V2 sequence C-strand (residues 165–186) were associated with viral escape antibody response.[32] However, we did not consistently found these changes in our patients.

The V3 region has been reported to be critical for env-mediated HIV-1-cell fusion.[23] However, we did not find any significant mutations associated with viral replication in the V3 region. Consistent with the finding of Morikita et al.,[33] we also found that the V3 region was quite homogeneous, whereas the V1 and V2 regions were relatively heterogeneous. The V3 region has been thought to be the crucial determinant of CCR5 co-receptor usage; when the V1/V2 and V3 domains were present, there was increased HIV viral infectivity.[34] Shalekoff et al.[35] found that the switch from CCR5 usage to CXCR4 usage in HIV-infected children was observed between 9 and 15 months. All our children were older than 15 months, so that we could speculate that most of them were using CXCR4. Therefore, with a crucial role in viral fusion activity and cellular tropism and a minor role in developing immune response, the V3 region could remain homogeneous without affecting viral replication.

Mutations in the V4 region including multiple point mutations, insertions, and deletions were detected in patients with high levels of HIV viral load. Sundravaradan et al.[36] have suggested that due to differences in sequences of V3 and V4 regions, patients with subtype C might encounter higher levels of HIV-1 replication, compared to those with subtype B, suggesting that sequences of the V3 and V4 regions might affect viral fitness and viral replication. In our study, the V3 region remained relatively homogeneous; mutations in the V4 region might be associated with increased viral replication. Taken together, our results suggested that HIV might acquire mutations in different regions of the *env* gene, including both conserved C2 and C3 and variable V1/V2 and V4 regions, in order to increase viral infectivity and replication capacity, while V3, which is required for fusion activity and cellular tropism, remains relatively homogeneous.

Limitation of the study

Our study is restricted to observations of changes of different regions of the env sequences. Since we have consumed all samples collected in a previous study, the numbers of patients for this study were limited. A more comprehensive functional study should be conducted in order to understand the association between the observed env variations and viral binding and immune-escape mutations for the particular AE subtype in Vietnam.

Conclusion

Our results suggest that env sequences of HIV are mutated in different regions, including both conserved C2 and C3 and variable V1/V2 and V4 regions in order to increase viral infectivity and replication capacity. Notably, the V3 region

remained relatively homogeneous, supporting the role of V3 for fusion activity and cellular tropism. Studying the genetic variability of the HIV virus in relation to viral replication and immunological status might be beneficial for prediction of disease progression and treatment responses during HIV infections.

ORCID iD

Linh Vu Phuong Dang (iD) https://orcid.org/0000-0001-6886-2713

References

1. VAAC. HIV prevention and control. 2016.
2. UNAIDS. The Global HIV/AIDS Epidemic. 2017.
3. Matte C, Zijenah LS, Lacaille J, et al. Mother-to-child human leukocyte antigen G concordance: no impact on the risk of vertical transmission of HIV-1. *AIDS* 2002; 16: 2491–2494.
4. Ngoufack MN, Nkenfou CN, Tiedeu BA, et al. CCR2 polymorphism and HIV: mutation in both mother and child is associated with higher transmission. *Int J Biochem Mol Biol* 2019; 10: 42–48.
5. Naidoo VL, Mann JK, Noble C, et al. Mother-to-child HIV transmission bottleneck selects for consensus virus with lower gag-protease-driven replication capacity. *J Virol*. Epub ahead of print 9 August DOI: 10.1128/JVI.00518-17.
6. Hemelaar J, Gouws E, Ghys PD, et al. Global trends in molecular epidemiology of HIV-1 during 2000–2007. *AIDS* 2011; 25: 679–689.
7. Roques P, Robertson DL, Souquiere S, et al. Phylogenetic analysis of 49 newly derived HIV-1 group O strains: high viral diversity but no group M-like subtype structure. *Virology* 2002; 302: 259–273.
8. Hemelaar J, Gouws E, Ghys PD, et al. Global and regional distribution of HIV-1 genetic subtypes and recombinants in 2004. *AIDS* 2006; 20: W13–W23.
9. Thomson MM and Nájera R. Molecular epidemiology of HIV-1 variants in the global AIDS pandemic: an update. *AIDS Rev* 2005; 7(4): 210–224.
10. Taylor BS, Sobieszczyk ME, McCutchan FE, et al. The challenge of HIV-1 subtype diversity. *N Engl J Med* 2008; 358: 1590–1602.
11. Lazaro E, Tram LT, Bellecave P, et al. Molecular characterization of HIV-1 CRF01_AE in Mekong Delta, Vietnam, and impact of T-cell epitope mutations on HLA recognition (ANRS 12159). *PLoS One* 2011; 6(10): e26244.
12. Liao H, Tee KK, Hase S, et al. Phylodynamic analysis of the dissemination of HIV-1 CRF01_AE in Vietnam. *Virology* 2009; 391: 51–56.
13. Sattentau QJ and Weiss RA. The CD4 antigen: physiological ligand and HIV receptor. *Cell* 1988; 52: 631–633.
14. Feng Y, Broder CC, Kennedy PE, et al. HIV-1 entry cofactor: functional cDNA cloning of a seven-transmembrane, G protein-coupled receptor. *Science* 1996; 272: 872–877.
15. Alkhatib G, Combadiere C, Broder CC, et al. CC CKR5: a RANTES, MIP-1alpha, MIP-1beta receptor as a fusion cofactor for macrophage-tropic HIV-1. *Science* 1996; 272: 1955–1958.
16. Arrildt KT, Joseph SB and Swanstrom R. The HIV-1 env protein: a coat of many colors. *Curr HIV/AIDS Rep* 2012; 9(1): 52–63.
17. Barroso H, Borrego P, Bartolo I, et al. Evolutionary and structural features of the C2, V3 and C3 envelope regions underlying the differences in HIV-1 and HIV-2 biology and infection. *PLoS One* 2011; 6: e14548.
18. Leonard CK, Spellman MW, Riddle L, et al. Assignment of intrachain disulfide bonds and characterization of potential glycosylation sites of the type 1 recombinant human immunodeficiency virus envelope glycoprotein (gp120) expressed in Chinese hamster ovary cells. *J Biol Chem* 1990; 265: 10373–10382.
19. Etemad B, Gonzalez OA, McDonough S, et al. Early infection HIV-1 envelope V1-V2 genotypes do not enhance binding or replication in cells expressing high levels of alpha4beta7 integrin. *J Acquir Immune Defic Syndr* 2013; 64: 249–253.
20. Toohey K, Wehrly K, Nishio J, et al. Human immunodeficiency virus envelope V1 and V2 regions influence replication efficiency in macrophages by affecting virus spread. *Virology* 1995; 213: 70–79.
21. Cao J, Sullivan N, Desjardin E, et al. Replication and neutralization of human immunodeficiency virus type 1 lacking the V1 and V2 variable loops of the gp120 envelope glycoprotein. *J Virol* 1997; 71(12): 9808–9812.
22. Johnson WE, Morgan J, Reitter J, et al. A replication-competent, neutralization-sensitive variant of simian immuno-deficiency

virus lacking 100 amino acids of envelope. *J Virol* 2002; 76(5): 2075–2086.

23. Saunders CJ, McCaffrey RA, Zharkikh I, et al. The V1, V2, and V3 regions of the human immunodeficiency virus type 1 envelope differentially affect the viral phenotype in an isolate-dependent manner. *J Virol* 2005; 79(14): 9069–9080.

24. Davis KL, Gray ES, Moore PL, et al. High titer HIV-1 V3-specific antibodies with broad reactivity but low neutralizing potency in acute infection and following vaccination. *Virology* 2009; 387: 414–426.

25. Lei L, Yang YR, Tran K, et al. The HIV-1 envelope glycoprotein C3/V4 region defines a prevalent neutralization epitope following immunization. *Cell Rep* 2019; 27: 586–598.e6.

26. Quan Y, Liang C, Brenner BG, et al. Multidrug-resistant variants of HIV type 1 (HIV-1) can exist in cells as defective quasispecies and be rescued by superinfection with other defective HIV-1 variants. *J Infect Dis* 2009; 200: 1479–1483.

27. Kwong PD, Wyatt R, Robinson J, et al. Structure of an HIV gp120 envelope glycoprotein in complex with the CD4 receptor and a neutralizing human antibody. *Nature* 1998; 393: 648–659.

28. Quan Y, Xu H, Kramer VG, et al. Identification of an env-defective HIV-1 mutant capable of spontaneous reversion to a wild-type phenotype in certain T-cell lines. *Virol J* 2014; 11: 177.

29. Pham HV, Ishizaki A, Nguyen LV, et al. Two-year outcome of first-line antiretroviral therapy among HIV-1 vertically-infected children in Hanoi, Vietnam. *Int J STD AIDS* 2015; 26(11): 821–830.

30. Samleerat T, Braibant M, Jourdain G, et al. Characteristics of HIV type 1 (HIV-1) glycoprotein 120 env sequences in mother-infant pairs infected with HIV-1 subtype CRF01_AE. *J Infect Dis* 2008; 198: 868–876.

31. Wu E, Du Y, Gao X, et al. V1 and V2 domains of HIV envelope contribute to CCR5 antagonist resistance. *J Virol* 2019; 93: e00050-e00519.

32. Trinh HV, Gohain N, Pham PT, et al. Humoral response to the HIV-1 envelope V2 region in a Thai early acute infection cohort. *Cells* 2019; 8.

33. Morikita T, Maeda Y, Fujii S, et al. The V1/V2 region of human immunodeficiency virus type 1 modulates the sensitivity to neutralization by soluble CD4 and cellular tropism. *AIDS Res Hum Retroviruses* 1997; 13: 1291–1299.

34. Foda M, Harada S and Maeda Y. Role of V3 independent domains on a dualtropic human immunodeficiency virus type 1 (HIV-1) envelope gp120 in CCR5 coreceptor utilization and viral infectivity. *Microbiol Immunol* 2001; 45(7): 521–530.

35. Shalekoff S, Gray GE and Tiemessen CT. Age-related changes in expression of CXCR4 and CCR5 on peripheral blood leukocytes from uninfected infants born to human immunodeficiency virus type 1-infected mothers. *Clin Diagn Lab Immunol* 2004; 11(1): 229–234.

36. Sundaravaradan V, Das SR, Ramakrishnan R, et al. Role of HIV-1 subtype C envelope V3 to V5 regions in viral entry, coreceptor utilization and replication efficiency in primary T-lymphocytes and monocyte-derived macrophages. *Virol J* 2007; 4: 126.