Emergence of a New HIV Type 1 CRF01_AE Variant in Guangxi, Southern China

Haiyan Zeng,¹ Zhiwu Sun,¹ Shujia Liang,² Lingnuo Li,¹ Yanyan Jiang,¹ Wei Liu,²
Binlian Sun,¹ Jingyun Li,³ and Rongge Yang¹

Abstract

The distribution of HIV-1 subtypes and genetic characterization of CRF01_AE in Guangxi, southern China were identified. The distribution of HIV-1 genotypes based on *gag*, *pol*, and partial *env* sequences ($n=349$) was as follows: CRF01_AE (66.5%), CRF08_BC (19.2%), CRF07_BC (7.2%), URF (4.6%), subtype B (1.7%), and subtype B' (0.9%). CRF01_AE predominated in all geographic regions and risk populations and there were multiple introductions of CRF01_AE strains in Guangxi. We found a peculiar CRF01_AE monophyletic lineage distinct from other CRF01_AE viruses, and we designated it "CRF01_AE-v" for convenience. CRF01_AE-v circulating in both heterosexuals and injecting drug users (IDUs) had accounted for 39.7% of CRF01_AE. It showed a selective advantage in the Guangxi population and formed its own characteristic compared with all the CRF01_AE references. Our results suggested that CRF01_AE-v was a new variant of CRF01_AE and it might lead to a new epidemic in Guangxi.

HETEROSEXUAL TRANSMISSION HAS become the most important transmission mode of HIV-1 in China since 2007, and CRF01_AE is a predominant subtype in the sexual risk population.^{1,2} Numerous studies reported that CRF01_AE had led a new epidemic in many major HIV-1 prevalent provinces and municipalities of China, and displayed a complex characteristic in these regions.¹⁻⁵ Guangxi province, which borders Vietnam in the south and Yunnan province in the west, is one of the provinces most affected by the HIV/AIDS epidemic in China. CRF01_AE and CRF08_BC were two major subtypes in Guangxi.⁶ Reports in recent years in Guangxi showed that CRF01_AE had dominated in the central city Nanning and the eastern city Liuzhou,⁷ and there may be more than one introduction of CRF01_AE in Guangxi.⁸ To gain a deeper understanding on the evolution of the HIV-1 epidemic and genetic characterization of CRF01_AE in Guangxi, we recruited 349 HIV/AIDS patients from cities all over Guangxi.

Subjects ($n=349$) were recruited between March 2009 and November 2010 from all the cities in Guangxi province, and the clinical and epidemiologic backgrounds of patients were collected through an interview. This study was approved by the committee for Human Research in the Center for Disease

Prevention and Control (CDC) of Guangxi province. The distribution of these subjects ($n=336$) is shown in Fig. 1, and the places of 13 subjects had not been recorded. The study subjects ($n=349$) included 224 heterosexuals, 11 homosexuals, 85 injecting drug users (IDUs), 1 mother-to-child transmission patients (MTCT), and 28 unknown. There were 250 men, 88 women, and 11 volunteers of unrecorded sex. Subjects had an age range of 18 to 82 years old (mean: 41.6 years). All 349 specimens were serologically found to be HIV-1 positive. No HIV-2 infections were detected. Baseline CD4⁺ T cell counts were obtained for 186 study subjects (mean: 378.56 cells/ μ l, range: 5–1036 cells/ μ l).

The 1.5-kb *gag*, 3.0-kb *pol*, and 1.1-kb partial *env* sequences were amplified from plasma RNA using reverse transcriptase polymerase chain reaction (RT-PCR). PCR products were subjected to DNA sequencing directly (all the primers for PCR and sequencing are available from the authors upon request). All 349 samples were amplified and sequenced successfully. Of these 349 samples, 236 were positive for *gag*, 227 were positive for *pol*, and 190 were positive for *env*. In addition, 77 samples were concordantly positive for all three genes, 72 for *gag* and *pol*, 39 for *gag* and *env*, 41 for *pol* and *env*; 48 for *gag* only, 37 for *pol* only, and 33 for *env* only. All nucleotide

¹Research Group of HIV Molecular Epidemiology and Virology, The State Key Laboratory of Virology, Wuhan Institute of Virology, Chinese Academy of Sciences, Wuhan, Hubei, The People's Republic of China.

²Guangxi Province Center for Disease Prevention and Control, Nanning, Guangxi, The People's Republic of China.

³Department of AIDS Research, The State Key Laboratory of Pathogen and Biosecurity, Beijing Institute of Microbiology and Epidemiology, Beijing, The People's Republic of China.

FIG. 1. Neighbor-joining tree of HIV-1 sequences from Guangxi based on the 1.5-kb HIV-1 *gag* region (HXB2: 792–2310 nt). SIVcpzUS is used as an outgroup. For clarity, only certain representative sequences are shown. The bootstrap values greater than 70 are shown at the corresponding nodes. Symbols represent different sex and risk groups: ● male heterosexual; ○ female heterosexual; ▲ male IDU; △ female IDU; ▼ male homosexual; ▽ female homosexual; ◆ male MTCT; □ unknown. Map of the People's Republic of China and the geographic distribution of studied HIV-1-infected cases in Guangxi are depicted at the upper right corner.

sequences obtained were screened using the BLAST program (National Center for Biotechnology Information, U.S.) to search for similarities to previously reported sequences in the databases and eliminate potential laboratory errors. All sequences were aligned with the Clustal X 2.0 program. Phylogenetic analyses were performed using the neighbor-joining and maximum likelihood methods implemented by the MEGA 5.0 program and intersubject distances were also calculated.⁹ The reliability of the topology of the tree was evaluated by bootstrap analyses with 1000 replicates. Bootscanning was used to assess the recombination breakpoints.¹⁰ Microsoft Excel was used to compare the intersubject average genetic distance between two groups with the nonpaired samples *t* test.⁶ *p* values represent the results of two-tailed tests. The VESPA program was used to detect signature patterns.¹¹ The resulting pattern consisted of all sites from the query sequence set that differed from what was found in over 50% of sequences in the reference set. The sequence logo was generated by WebLogo.¹²

The distribution of HIV-1 genotypes based on *gag*, *pol*, and partial *env* sequences ($n=349$) was as follows: CRF01_AE (232/349, 66.5%), CRF08_BC (67/349, 19.2%), CRF07_BC (25/349, 7.2%), URF (16/349, 4.6%), subtype B (6/349, 1.7%), and subtype B' (3/349, 0.9%). Bootscanning analysis of *gag/pol/env* sequences revealed that the unique recombinant strains of CRF01_AE/B, CRF01_AE/BC, CRF01_AE/CRF07_BC, CRF01_AE/CRF08_BC, and CRF01_AE/C, BC presented in Guangxi (data not shown). CRF01_AE predominated in heterosexuals (171/224, 76.3%), homosexuals (6/11, 54.5%), and IDUs (40/87, 46.0%). CRF08_BC was still found more in IDUs (33/67, 49.3%) than in heterosexuals (25/67, 37.3%), while

CRF07_BC was found more in heterosexuals (15/25, 60%) than in IDUs (6/25, 24%). The geographic distribution of the epidemic of CRFs in HIV-1-infected residents of Guangxi showed that CRF01_AE had spread all over Guangxi, CRF08_BC was dispersed in most cities of Guangxi, and CRF07_BC was mainly found in the center of and in eastern Guangxi.

Three clusters designated as clusters I, II, and III were observed among CRF01_AE sequences in phylogenetic tree analysis based on the *gag*, *pol*, and partial *env* sequences. We only presented the neighbor-joining tree based on *gag* sequences in Fig. 1. Cluster I, along with some strains previously isolated in Guangxi, grouped with some strains prevalent in Vietnam. This proved that there was a close epidemiologic relationship between the CRF01_AE strains found in Guangxi and those found in Vietnam.¹³ Cluster II including two strains isolated in Guangxi before a distinct cluster without any international reference sequences was formed. Some strains, along with the reference strains from Fujian and Jiangsu provinces of China, gathered into cluster III. These results implied that there was more than one introduction of CRF01_AE in Guangxi. Compared to cluster I, cluster II, which diverged from other CRF01_AE references, gathered into a similar monophyletic lineage with a higher bootstrap value and a tighter topology structure. The strains in cluster II were different from those in cluster I, which were also epidemic in Guangxi. To gain a deeper understanding of the genetic characteristics of CRF01_AE in Guangxi, we analyzed the peculiar cluster II comparing it with cluster I as follows.

FIG. 2. Distribution of the signature amino acids of Guangxi CRF01_AE cluster II in the genome of HIV-1. As shown at the top, the genomic regions we amplified successfully are marked in blue. The amino acid mutations in Guangxi cluster II comparing with the CRF01_AE reference set are depicted in several black boxes (locations are defined by HXB2). The amino acid alterations of the Guangxi CRF01_AE cluster II are mainly found in p17, p6, *gag-pol* fusion TF protein, V2 loop, and C3 region of gp120. Sequence logos for the *env* V3/C3 region (HXB2: 7095–7397) of the CRF01_AE reference set, Guangxi clusters II and I, are showed at the bottom. The consensus sequence and relative frequency of amino acids can be read in the sequence logos. The GPGQ core of the V3 loop is marked with a red square. Boxed amino acids at positions 293 (293L), 337 (337N), 339 (339F), 346 (346A), 364 (364H), and 388 (388S) are the signature amino acids for Guangxi cluster II.

Cluster I ($n=119$) included 98 heterosexuals, 14 IDUs, one homosexual, and six unknown subjects. Cluster II ($n=92$) included 60 heterosexuals, 23 IDUs, one homosexual, and eight unknown subjects. Both cluster I and cluster II were circulating in heterosexuals and IDUs. The comparison of the distribution between these two clusters in the main cities showed that both cluster I and cluster II had a high distribution in central and western Guangxi, whereas the strains in cluster II were more common in Guangxi inland such as Guigang and Laibang cities.

We also calculated the intersubject nucleotide diversity of clusters I and II. For *gag*, *pol*, and partial *env*, the intersubject variations were $6.20 \pm 0.78\%$, $4.1 \pm 0.39\%$, and $16.9 \pm 1.38\%$ for cluster I and $5.20 \pm 0.59\%$, $3.0 \pm 0.20\%$, and $14.7 \pm 1.38\%$ for cluster II, which were significantly different for the sequences of all three genes ($0.02 < p < 0.05$). Cluster II had a lower genetic variation than cluster I. As an epidemic spreads rapidly through a population, the evolutionary rate of HIV-1 is relatively low and the diversity of virus is limited.¹⁴ These results suggested that cluster II had a selective advantage in the Guangxi population compared with cluster I.

A unique amino acid signature pattern was detected in cluster II by comparing an alignment with all CRF01_AE reference sequences from the Los Alamos HIV database (<http://www.hiv.lanl.gov/content/index>) using VESPA. The signature pattern of cluster II consisted of 33 noncontiguous amino acids (see Fig. 2 for the positions of these amino acid alterations). Only 7 of 33 signature amino acids of cluster II were found in cluster I. This result suggested that cluster II and I had a close phylogenetic relationship, but they were different. The mutations of T211I in the reverse transcriptase and A34V in the integrase region have not been reported to be relevant to drug resistance. The N-linked glycosylation site (N339) disappears as a result of the alteration of N339F in the C3 region of gp120 (Fig. 2). N339 was reported to likely play an indirect role in the binding of the broadly neutralizing human monoclonal antibody 2G12.¹⁵ The effects of other specific amino acid alteration are still unknown.

We designate cluster II "CRF01_AE-v" for convenience. CRF01_AE-v may be a new variant of CRF01_AE and it is the product of a highly variable CRF01_AE strain that has spread rapidly in Guangxi. The following three lines of evidence strongly support this hypothesis: (1) CRF01_AE-v, which diverged from other CRF01_AE strains, forms a distinct monophyletic lineage. (2) CRF01_AE-v, which has accounted for 39.7% (92/232) of CRF01_AE, has a lower genetic variation than cluster I, showing a selective advantage in the Guangxi population. CRF01_AE-v is circulating in heterosexuals and IDUs; this phenomenon shows that the influence of the human factor in the epidemic cannot be excluded. (3) A unique amino acid signature pattern is detected in CRF01_AE-v strains relative to the entire CRF01_AE reference set (Fig. 2). This implies that CRF01_AE-v has formed its own characteristic. CRF01_AE-v with these features may facilitate the spread of the virus. In conclusion, our results suggest that CRF01_AE-v is a new variant of CRF01_AE and it may lead to a new epidemic in Guangxi.

This is the first report of a province-wide HIV-1 molecular epidemiological study in Guangxi. Our study promotes a deeper understanding of the evolution of the HIV-1 epi-

demic and the genetic characteristics of CRF01_AE strains in Guangxi. Further studies in Guangxi and neighboring regions will be useful to determine whether this new CRF01_AE variant will be able to obtain an advantage in the spread to other places. We recommend that the difference between CRF01_AE and CRF01_AE-v should be taken into consideration in the investigation of the epidemic and in vaccine design.

Acknowledgments

The authors would like to thank Yue Li and Wei Zhang for helpful suggestions regarding data interpretation. This study was supported by the Key National Science and Technology Program in the 11th Five-Year Period (Grant 2008ZX10001-004). The GenBank accession numbers of the nucleotide sequences reported in this article are JQ028147~JQ028668.

Author Disclosure Statement

No competing financial interests exist.

References

1. Zhang Y, Lu L, Ba L, *et al.*: Dominance of HIV-1 subtype CRF01_AE in sexually acquired cases leads to a new epidemic in Yunnan province of China. *PLoS Med* 2006; 3(11):e443.
2. Guo D, Ding N, Xu Y, *et al.*: Near full-length genome characterization of an HIV-1 CRF01_AE strain in Jiangsu, China: Evidence of two independent introductions from Fujian. *AIDS Res Hum Retroviruses* 2009;25(6):619–623.
3. Han X, Dai D, Zhao B, *et al.*: Genetic and epidemiologic characterization of HIV-1 infection in Liaoning Province, China. *J Acquir Immune Defic Syndr* 2010;53(Suppl 1):S27–33.
4. Deng W, Fu P, Bao L, *et al.*: Molecular epidemiological tracing of HIV-1 outbreaks in Hainan island of southern China. *AIDS* 2009;23(8):977–985.
5. Zhong P, Pan Q, Ning Z, *et al.*: Genetic diversity and drug resistance of human immunodeficiency virus type 1 (HIV-1) strains circulating in Shanghai. *AIDS Res Hum Retroviruses* 2007;23(7):847–856.
6. Laeyendecker O, Zhang GW, Quinn TC, *et al.*: Molecular epidemiology of HIV-1 subtypes in southern China. *J Acquir Immune Defic Syndr* 2005;38(3):356–362.
7. Su Q, Liang H, Cen P, Bi ZY, and Zhou P: HIV-1 subtypes based on the *pol* gene and drug resistance mutations among antiretroviral-naïve patients from Guangxi, Southern China. *AIDS Res Hum Retroviruses* 2011 [Epub ahead of print]; DOI: 10.1089/aid.2011.0246.
8. Li L, Liang S, Chen L, *et al.*: Genetic characterization of 13 subtype CRF01_AE near full-length genomes in Guangxi, China. *AIDS Res Hum Retroviruses* 2010;26(6):699–704.
9. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, and Kumar S: MEGA5: Molecular Evolutionary Genetics Analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol* 2011;28(10):2731–2739.
10. Lole KS, Bollinger RC, Paranjape RS, *et al.*: Full-length human immunodeficiency virus type 1 genomes from subtype C-infected seroconverters in India, with evidence of intersubtype recombination. *J Virol* 1999;73(1):152–160.
11. Korber B and Myers G: Signature pattern analysis: A method for assessing viral sequence relatedness. *AIDS Res Hum Retroviruses* 1992;8(9):1549–1560.

12. Crooks GE, Hon G, Chandonia JM, and Brenner SE: WebLogo: A sequence logo generator. *Genome Res* 2004; 14(6):1188–1190.
13. Kato K, Kusagawa S, Motomura K, *et al.*: Closely related HIV-1 CRF01_AE variant among injecting drug users in northern Vietnam: Evidence of HIV spread across the Vietnam-China border. *AIDS Res Hum Retroviruses* 2001; 17(2):113–123.
14. Maljkovic Berry I, Ribeiro R, Kothari M, *et al.*: Unequal evolutionary rates in the human immunodeficiency virus type 1 (HIV-1) pandemic: The evolutionary rate of HIV-1 slows down when the epidemic rate increases. *J Virol* 2007; 81(19):10625–10635.
15. Chaillon A, Braibant M, Moreau T, *et al.*: The V1V2 domain and an N-linked glycosylation site in the V3 loop of the

HIV-1 envelope glycoprotein modulate neutralization sensitivity to the human broadly neutralizing antibody 2G12. *J Virol* 2011;85(7):3642–3648.

Address correspondence to:

Rongge Yang
Research Group of HIV Molecular Epidemiology and Virology
The State Key Laboratory of Virology
Wuhan Institute of Virology
Chinese Academy of Sciences
Xiaohongshan 44
Wuhan 430071
People's Republic of China
E-mail: ryang@wh.iov.cn