Recombination Form and Epidemiology of HIV-1 Unique Recombinant Strains Identified in Yunnan, China

Lin Li¹, Lili Chen¹,³, Shaomin Yang², Tianyi Li¹, Jianjian Li², Yongjian Liu¹, Lei Jia¹, Bihui Yang², Zuoyi Bao¹, Hanping Li¹, Xiaolin Wang¹, Daomin Zhuang¹, Siyang Liu¹, Jingyun Li¹*¹

Abstract
Several studies identified HIV-1 recombination in some distinct areas in Yunnan, China. However, no comprehensive studies had been fulfilled in the whole province up to now. To illustrate the epidemiology and recombination form of Unique Recombinant Forms (URFs) circulating in Yunnan, 788 HIV-1 positive individuals residing in 15 prefectures of Yunnan were randomly enrolled into the study. Full-length gag and pol genes were amplified and sequenced. Maximum likelihood tree was constructed for phylogenetic analysis. Recombinant breakpoints and genomic schematics were identified with online software jPHMM. 63 (10.2%) unique recombinant strains were identified from 617 strains with subtypes. The URFs distributed significantly differently among prefectures (Pearson chi-square test, P<0.05). IDUs contained more URFs than sexual transmitted population (Pearson chi-square test, P<0.05). Two main recombinant forms were identified by considering the presence of CRF01_AE segments in full length gag-pol genes, which were B/C and B'/C/CRF01-AE recombinants. Three clusters were identified in the ML tree which contained more than three sequences and supported by high bootstrap values. One CRF was identified. Many of URFs contained identical breakpoints. The results will contribute to our understanding on HIV recombination and provide clues to the identification of potential CRFs in China.

Introduction
Human immunodeficiency viruses (HIV) were characterized with high level of genetic variation. Recombination is one of main mechanisms of HIV diversity. When a cell that was dually infected by two different viruses produced progeny virions with genomic RNAs from each virus, strand-switching would take place during the next round of reverse transcription [1,2,3]. As the results, recombination would happen and unique variants with genome from two distinct parental viruses would be produced. Substantial studies showed that recombination contributed to viral escaping mutations under immune pressure, viral fitness and emergence of viral drug-resistance [4]. Up to now, 52 circulating recombinant forms (CRFs) (http://www.hiv.lanl.gov/content/hiv-db/CRFs/CRFs.html) have been reported, which is responsible for more than 20% HIV infections in the global AIDS epidemic [5]. Many studies demonstrated that mosaic strains were arising frequently, especially in populations with multiple subtypes circulating [6], surveillance on the emergence and epidemic of HIV-1 unique recombinant forms will be helpful for the prediction of CRFs and HIV epidemic in different area.

Yunnan, a southwestern province of China that shares a border with the famous heroin-producing area of Myanmar (Burma), was considered as the epicenter of China. The first HIV-1 epidemic that occurred among injecting drug users (IDUs) in 1989 in China was reported in Yunnan [7,8,9]. Now it is still the area with the most severely HIV epidemic in China with highest number of new reported cases. Multiple HIV-1 genotypes, including B, C, CRF01_AE, CRF07_BC, and CRF08_BC, are all currently circulating in the area [10], supposing the possible emergence of a new recombination. CRF07_BC and CRF08_BC, the dominant CRF strains spreading in China now, were believed to arose in Yunnan in 1990s and spread to other areas [11,12]. In recent years, the transmitting routes and prevalence of different HIV subtypes or CRFs have changed dramatically in Yunnan [13]. Accordingly, URFs originated from prevalent subtypes in same population would change. Recently, extensive recombination of HIV was identified in Myanmar, which is close to Yunnan province [14]. All of these reasons urgent the surveillance of HIV recombination in Yunnan, China. In this study, we characterized the genomic schematic and illustrated the epidemic of HIV URFs in whole Yunnan province basing on full length gag-pol gene sequences.

Materials and Methods
Study subjects and specimens
HIV-positive individuals were randomly recruited from the list of Yunnan AIDS care center, with written informed consents. The ratios of enrolled subjects from different prefectures were determined according to the reported cases of prefectures to make
them representatively. The epidemiologic background was collected through specific epidemiological investigation by trained interviewers. Peripheral blood was sampled from Jan. 2008 to Dec. 2009, and plasma was separated and stored in −80°C freezer. The Ethical Review Board, Science and Technology Supervisory Committee at the Beijing Institute of Microbiology and Epidemiology approved the study.

HIV-1 RNA extraction, amplification and sequencing

Viral RNA was extracted from 500 µl HIV-1 positive plasma specimens with high pure viral RNA kit (Roche, USA). Viral full length gag (from 763 to 2400 according to the HXB2 calibrator) and pol (from 2068 to 5221 according to the HXB2 calibrator) genes were amplified separately with more than 300 bp overlapping region using reverse transcriptional nest PCR as described before [15]. Positive PCR products were sequenced by Huaada genomics company (China) with a variety of internal specific primers (available on request) after being purified. For variant containing breakpoint in the overlapping regions of gag and pol genes, 5' half genome was amplified and sequenced again using methods provide in previous paper [16]. The nearly full-length genomes were reverse-transcribed and amplified in two halves with 1 kb overlapping regions as described before [16].

Edit, Assemble, Genotyping and Phylogenetic Analysis of HIV-1 Sequences

All of the sequenced fragments were edited and assembled as described before [16]. To check for potential contamination, the sequences obtained were compared to all known sequences in the HIV database by a Basic Local Alignment Search Tool (BLAST) search (http://hiv-web.lanl.gov/content/index). HIV genotype was determined using the national center for biotechnology information viral genotyping tool (http://www.ncbi.nlm.nih.gov/projects/genotyping/formpage.cgi) by combining both gene regions from same isolate and further confirmed by phylogenetic analysis with reference sequences using the Neighbor-Joining method in MEGA5.0 software and Maximum Likelihood method in PhyML software [17]. To fulfill the phylogenetic analysis, full length gag and pol genes derived from same variant were assembled together and aligned with reference sequences representing subtypes A-D, F-H, J, K, CRF01_AE, CRF07_BC, CRF08_BC and CRF31_BC (http://www.hiv.lanl.gov). ML tree was constructed using GTR + gamma model of nucleotide substitution which was selected with jModelTest software [17,18]. SPR was fulfilled to explore possible evolution relationship. To do this, all gag-pol genes as described in the paper were downloaded from Los Alamos database and analyzed together with our sequences. Since CRF01_AE, CRF07_BC, CRF08_BC and subtype B' were the main prevalent strains in Yunnan province, it is not surprising for us to find that all genomes of the URFs were comprised with them (Table 2). Strains composed of subtype B' and C in the gag-pol regions (47 cases) were dominant in the URFs, which were responsible for 74.6% of total URFs. Subtype B' and C were believed to be introduced into Yunnan province and responsible for early HIV epidemic, especially in the IDUs [20,21]. CRF01_AE was found to lead HIV epidemic in heterosexual transmitted population in recent years [15]. That may be the reason why most of URFs with CRF01_AE as genomic backbone were found in heterosexual transmitted population (75%). The distribution of URFs with different backbones was significantly different in IDUs and heterosexual transmitted population. Comparing to the data published by Yang et al in 2002, the composition of gag-pol genes also changed. The ratios of HIV-1 variants containing CRF01_AE segment in gag-pol genes increased from 8.3% in 2002 in Yang’s paper to 25.4% in 2009 in this manuscript, which strongly suggested that the recombination form of URFs might be changing.

Phylogenetic analysis of URFs

In the maximum likelihood tree constructed with full-length gag-pol genes sequenced in this study, all of unique recombinant strains placed among the subtype reference sequences, with three clusters containing more than 3 sequences with high branch supports in the ML tree. (Figure 1) With detailed URF sequence data presented by Yang et al in 2002, phylogenetic analysis was further fulfilled to explore possible evolution relationship. To do this, all gag-pol genes from URFs reported in this study were aligned with 12 URF sequences identified in Yang’s paper and ML tree was constructed basing on 2.6 Kb gag-RT genes as described in the materials and methods. Similar topology was observed as the ML tree constructed with only our sequences. DH003, DH008 and DH012 from Yang’s lab were found locating in Cluster III. To identify more HIV-1 variants with similar recombinant form, all of Asian URFs (57 strains) containing subtype B, C and CRF01_AE gene segments in gag-pol regions were downloaded from Los Alamos database and analyzed together with our sequences. Phylogenetic tree showed that YNRL9607, YNRL9613 grouped into cluster II. Strain 341 grouped into cluster III.
Analysis of recombinant breakpoint and identification of a novel CRF

Recombinant breakpoints analysis basing on full length gag-pol sequences showed that 56 of 63 URFs contained same breakpoints with at least one of other strains (Figure 2). Similar genomic schematic structures among sequences in the same cluster were found. In cluster I, 813 and 814 contained same breakpoints, while 816, 822 and 823 shared same breakpoint. All of those 5 strains were from same area, it is hard to exclude epidemiological linkage. In cluster II, 515 and 523 contained similar breakpoints in pol gene, while 523 had a short 230 bp length subtype B 9 gene segment inserting in gag region. The other sequences locating in cluster II contained longer subtype B 9 sequences in pol gene. 1444 and 2347 contained same genomic schematic and might be potential CRF. All of sequences locating in cluster III contained same recombinant breakpoints (1229 and 1837 according to HXB2 calculator) with subtype B’ gene segments inserted in gag region. The other sequences locating in cluster II contained longer subtype B’ sequences in pol gene, 1444 and 2347 contained same genomic schematic and might be potential CRF. All of sequences locating in cluster III contained same recombinant breakpoints (1229 and 1837 according to HXB2 calculator) with subtype B’ gene segments inserted in gag region. Three strains (DH003, DH012 and DH015) in professor Yang’s paper contained same breakpoints as strains in cluster III identified in this study. Epidemiological background investigations excluded linkage among those participants (Table 3) and suggested that those strains might represent a new CRF. All full length gag-pol sequences of URFs containing B’, C or CRF01_AE segment from Asian were obtained from database and submitted to breakpoints analysis. The results showed that strain 341 had similar breakpoint as 1434, 1439, 2315, 2327 strains, and also the strain DH003 from Yang’s paper. To determine the full length genomic schematic, 3’ half genomic sequences of 1439 strain identified in this study were obtained as described in the materials and methods. The jpHMM analysis showed that 1439 and 341 strains displayed same genomic structure (Figure 2). In the chimera, a subtype B’ segment was inserted into the backbone of subtype C, with breakpoints corresponding to HXB2 nucleotide positions 1229 and 1837 approximately. According to the criteria for designation of a new CRF, those 6 strains constitute a new CRF identified in the worldwide HIV-1 pandemic.

Discussion

High heterogeneity of HIV was always observed in infected individuals, which made it difficult for the amplification and sequencing of long gene segments. In most of previous studies, the determination of HIV subtypes prevalent in Yunnan province was always based on short gene segment, which could not conclude HIV-1 subtypes and identify recombination exactly. Long gene segment, even full length genomic sequence, is now considered more suitable for determining HIV subtypes. Many studies obtained long gene segment through assembling many short segments, however, it was hard to get gene segment originated from same variants due to the existence of quasispecies. In this study, we analyzed HIV genomic schematic structure basing on full length gag (1584 bp according to HXBII calculator) and pol (3147 bp according to HXBII calculator) genes, which make it more exactly for identification of recombinant strains.

HIV epidemic in Yunnan province was characterized with multiple strains prevalent simultaneously. The subtype constitution in Yunnan has always been changing. Accordingly, epidemiology of URFs would change. In 2002, Yang R et al. found that URFs accounted for 70% of HIV prevalent in IDUs in Dehong prefecture [22]. In this more comprehensive study basing on long gene segments, highest prevalence of URFs in IDU population was also observed. However, we found that the prevalence of URFs in IDUs in Dehong prefecture was 35% (14/40). Due to

---

Table 1. Distribution of URFs in different prefectures and transmitting routes in Yunnan.

| Prefectures and transmitting routes | URFs | Cases with subtypes | Prevalence (%) |
|-----------------------------------|------|---------------------|----------------|
| Prefectures                       |      |                     |                |
| Dehong                            | 29   | 94                  | 30.9           |
| Dali                              | 5    | 19                  | 26.3           |
| Kunming                           | 9    | 61                  | 14.8           |
| Zhaotong                          | 2    | 15                  | 13.3           |
| Puer                              | 4    | 38                  | 10.5           |
| Honghe                            | 4    | 41                  | 9.8            |
| Baoshan                           | 2    | 21                  | 9.5            |
| Xishuangbannan                    | 2    | 91                  | 2.2            |
| Lincang                           | 3    | 186                 | 1.6            |
| Other prefectures                 | 0    | 25                  | 0.0            |
| Unknown                           | 3    | 26                  | 11.5           |
| Transmitting routes               |      |                     |                |
| Sexual transmission               | 32   | 407                 | 7.9            |
| IDUs                              | 20   | 129                 | 15.5           |
| Others                            | 11   | 81                  | 14.3           |
| Total                             | 63   | 617                 | 10.2           |

doi:10.1371/journal.pone.0046777.t001

Table 2. Distribution of URFs of different patterns in transmitting routes.

| Patterns of URFs    | Transmitting routes | Total |
|---------------------|---------------------|-------|
|                     | Sexual transmission | IDUs  | Others |
| B’/C                | 20 (31.7%)          | 19 (30.2%) | 8 (12.7%) | 47 (74.6%) |
| B’/C/CRF01_AE       | 12 (19.0%)          | 11 (1.6%)  | 3 (4.8%)  | 16 (25.4%) |
| Total               | 32 (50.8%)          | 20 (31.7%) | 11 (17.5%) | 63 (100%)  |

doi:10.1371/journal.pone.0046777.t002
different genomic regions were adopted in determining viral subtypes, it is kind of hard to compare both results, but our data supposed the possible change of URFs prevalence in Yunnan. The recombinant forms of URFs were also in changing. In previous studies, URFs containing genomes composed of subtype B and C were dominant [22,23]. In this study, although recombinant forms of URFs were also in changing. In previous studies, URFs containing genomes composed of subtype B' and C were dominant [22,23]. In this study, although recombinant forms of URFs were also in changing.
strains originated from subtype B and C segments were still dominant in URFs, more URFs with CRF01_AE as the backbone were identified. CRF08_BC strains were dominating HIV epidemic in Yunnan province, which was responsible for more than 50% HIV cases (data not shown). Many URFs containing CRF08_BC breakpoints were identified in this study. So it was
necessary to pay attention to URFs originated from CRF08_BC which may become a CRF. CRF01_AE was believed to be introduced into Yunnan heterosexual transmitted population in few years [13] and leading a new epidemic in Yunnan. Identifications of URFs originated from it strongly suggested that the subtype is spreading quickly and need to be put more attention in the surveillance of HIV in the area.

In this study, at least three clusters of sequences with similar recombinant sites were identified. One CRF was identified with sequences obtained from HIV Los Alomas database and Yang’s paper [22]. Initially, subtype B’ was believed to be main subtypes in IDUs in Yunnan [20,21,24,25]. However, CRF07_BC and CRF08_BC strains spread quickly after their formation and became dominant in Yunnan and many other areas in China [11]. Comparing to their parental strains, CRF07_BC and CRF08_BC showed more superiority of spreading in Chinese population. Since many new URFs originated from now prevalent strains were observed, there is the possibility of new CRFs displaced them and become dominant. Further studies were necessary to explore the biological characterizations of URFs and social factors related to their epidemic, which will provide predictive information on HIV in Yunnan and even China.

Acknowledgments

We also want to show our thanks to Dr. Feng Gao in Duke Human Vaccine Institute, Duke University Medical Center for providing the technical support in reverse transcription and nest-PCR.

Author Contributions

Conceived and designed the experiments: LL, Jingyun Li. Performed the experiments: LL, LC, ZB, HL, XW, DZ, SL. Analyzed the data: LL, TL, LJ. Wrote the paper: LL, Jingyun Li, LJ. Participated in the collections and epidemiological investigation of patients: SY, Jianjian Li BY.

References

1. Goodrich DW, Duesberg PH (1990) Retroviral recombination during reverse transcription. Proc Natl Acad Sci U S A 87: 2052–2056.
2. Hu WS, Teman HM (1990) Retroviral recombination and reverse transcription. Science 250: 1227–1233.
3. Stuhlmann H, Berg P (1992) Homologous recombination of copackaged retrovirus RNAs during reverse transcription. J Virol 66: 2378–2388.
4. Sagar M, Wu X, Lee S, Overbaugh J (2006) Human immunodeficiency virus type1 V1–V2 envelope loop sequences expand and add glycosylation sites over the course of infection, and these modifications affect antibody neutralization sensitivity. J Virol 80: 9586–9598.
5. Hemelaar J, Gouws E, Ghys PD, Osmanov S (2011) Global trends in molecular epidemiology of HIV-1 during 2000–2007. AIDS 25: 679–689.
6. Takebe Y, Motomura K, Tatsumi M, Lwin HH, Zaw M, et al. (2003) High prevalence of diverse forms of HIV-1 inteotype recombinants in Central Myam: geographical hot spot of extensive recombination. AIDS 17: 2077–2087.
7. Zheng X, Tian C, Choi KH, Zhang J, Cheng H, et al. (1994) Injecting drug use and HIV infection in southwest China. AIDS 8: 1141–1147.
8. Xia M, Kreis JK, Holmes KK (1994) Risk factors for HIV infection among drug users in Yunnan province, China: association with intravenous drug use and protective effect of boiling reusable needles and syringes. AIDS 8: 1701–1706.
9. Sun X, Nan J, Guo Q (1994) AIDS and HIV infection in China. AIDS 8 Suppl 2: S33–59.
10. Tu YQ, Wang MJ, Yao J, Zhu XM, Pan PL, et al. (2009) Human immunodeficiency virus-1 genotype drug resistance among volunteer blood donors in Yunnan, China. Transfusion 49: 1051–1052.
11. Piyaniruip S, McCutchan FE, Carr JK, Sanders-Buell E, Liu W, et al. (2004) A recent outbreak of human immunodeficiency virus type 1 infection in southern China was initiated by two highly homogeneous, geographically separated strains, circulating recombinant form AE, and a novel BC recombinant. J Virol 74: 11280–11290.
12. Su L, Graf M, Zhang Y, von Briesen H, Xing H, et al. (2000) Characterization of a virtually full-length human immunodeficiency virus type 1 genome of a prevalent intersubtype (C/B') recombinant strain in China. J Virol 74: 11367–11376.
13. Zhang Y, Lu L, Bo L, Liu L, Yang L, et al. (2006) Dominance of HIV-1 subtype CRF01_AE in sexually acquired cases leads to a new epidemic in Yunnan province of China. PLoS Med 3: e443.
14. Pang W, Zhang C, Duo L, Zhou YH, Yao ZH, et al. (2012) Extensive and complex HIV-1 recombination between B’ and CRF01_AE among IDUs in Northern Myanmar. AIDS.
15. Li L, Li X, Li H, Chen L, Wang Z, et al. (2011) High genetic diversity of HIV-1 was found in men who have sex with men in Shijiazhuang, China. Infect Genet Evol 11: 1487–1492.
16. Li L, Liang S, Chen L, Liu W, Li H, et al. (2010) Genetic characterization of 13 subtype CRF01_AE near full-length genomes in Guangxi, China. AIDS Res Hum Retroviruses 26: 699–704.
17. Guindon S, Gascuel O (2003) A simple, fast, and accurate algorithm to estimate large phylogeny by maximum likelihood. Syst Biol 52: 191–199.
18. Posada D (2008) jModelTest: phylogenetic model averaging. Mol Biol Evol 25: 1253–1256.
19. Zhang M, Schulz AK, Calef C, Kuiken C, Leitner T, et al. (2006) jpHHMM at GORBICS: a web server to detect genomic recombinations in HIV-1. Nucleic Acids Res 34: W463–465.
20. Graf M, Shao Y, Zhao Q, Seidl T, Kostler J, et al. (1998) Cloning and characterization of a virtually full-length HIV type 1 genome from a subtype B’-Thai strain representing the most prevalent B-clade isolate in China. Nucleic Acids Res 35: W462–463.
21. Yu CC, Tian C, Hu DJ, Kai M, Doudens T, et al. (1993) HIV-1 subtype C in China. Lancet 343: 1051–1052.
22. Yang R, Xia X, Kusagawa S, Zhang C, Ben K, et al. (2002) On-going generation of multiple forms of HIV-1 intersubtype recombinants in the Yunnan Province of China. AIDS 16: 1401–1407.
23. Chen L, Yang S, Li J, Liu Y, Wang Z, et al. (2011) Near full-genome genetic characterization of a novel HIV type 1 CRF07_BC/CRF08_BC recombinant strain from Yunnan, China. AIDS Res Hum Retroviruses 27: 693–699.

Table 3. Information on subjects harboring novel B’/C recombinants in Yunnan China.

| Patients | Age | Gender | Country | Year of fist positive HIV test | Year of collection | Risk factor | Viral Load | CD4 cell count | Accession number |
|----------|-----|--------|---------|--------------------------------|-------------------|------------|-----------|--------------|-----------------|
| DH003*   | -   | CN     | -       | 2000–2001                      | IDU               | -          | -         | AB078705     |
| 1434     | 3   | Female | CN      | 2008                           | 2009              | MCT        | 109361    | 884          | JQ898195, JQ898254 |
| 1439     | 30  | Male   | CN      | 2008                           | 2009              | IDU        | 96639     | 379          | JX679207        |
| 2315     | 45  | Male   | CN      | 2007                           | 2009              | IDU        | 49795     | 276          | JQ898203, JQ898262 |
| 2327     | 9   | Female | CN      | 2007                           | 2009              | MCT        | 1361      | 142          | JQ898207, JQ898265 |
| 341#     | 35  | Male   | CN      | 2007                           | 2009              | IDU        | 17963     | 319          | HM776939        |

*: DH003 strain was identified by Yang et al. in 2002. [22].

#: 341 strain was downloaded from HIV Los Alomas database which was also published by our lab. [26].

\[ \text{doi:10.1371/journal.pone.0046777.t003} \]
24. Wagner R, Denl L, Teeuwen V, Henney J, Yiming S, et al. (1996) A recombinant HIV-1 virus-like particle vaccine: from concepts to a field study. Antimicrob Chemother 48: 68–83.

25. Weniger BG, Takebe Y, Ou CY, Yamazaki S (1994) The molecular epidemiology of HIV in Asia. AIDS 8 Suppl 2: S13–28.

26. Liu Y, Li L, Yang S, Bao Z, Li H, et al. (2011) Identification and characterization of two new HIV type 1 Unique (B/C) recombinant forms in China. AIDS Res Hum Retroviruses 27: 445–451.