

Multiple Independent Introductions of HIV-1 CRF01_AE Identified in China: What Are the Implications for Prevention?

Yassir F. Abubakar¹*, Zhefeng Meng¹*, Xiaoyan Zhang^{1,2}, Jianqing Xu^{1,2*}

1 Shanghai Public Health Clinical Center and Institutes of Biomedical Sciences, Key Laboratory of Medical Molecular Virology of Ministry of Education/Health at Shanghai Medical College, Fudan University, Shanghai, China, **2** State Key Laboratory of Infectious Disease Control and Prevention (CDC), Beijing, China

Abstract

Background: HIV-1 CRF01_AE accounts for an important fraction of HIV infections in Asia including China, but little is known about the phylogenetic and evolutionary history of this CRF (circulating recombinant form). In the current study, we collected a large number of 1,957 CRF01_AE gag p17 sequences with known sampling year (1990-2010) from 5 global regions representing 15 countries to better understand the phylogenetic relationships and epidemic history of CRF01_AE strains in China.

Methodology/Principal Findings: CRF01_AE gag p17 sequences were retrieved from public databases to explore phylogenetic relationships and phylogeographic dynamics of CRF01_AE in Asia by using maximum-likelihood phylogenetics and Bayesian coalescent-based analyses. We found close phylogenetic relationships between sequences from Thailand, Vietnam and China. Moreover, at least 5 independent introductions and 5 independent autochthonous clades of CRF01_AE, which descended from Thailand or Vietnam were identified in China from 1991 through 2003.

Conclusion/Significance: The current study not only defines the migration of CRF01_AE clades to/in Asia, but also demonstrates the criticalness of identifying the circulating strains in the population for the development of vaccine and microbicides.

Citation: Abubakar YF, Meng Z, Zhang X, Xu J (2013) Multiple Independent Introductions of HIV-1 CRF01_AE Identified in China: What Are the Implications for Prevention? PLoS ONE 8(11): e80487. doi:10.1371/journal.pone.0080487

Editor: Wing-Cheong Yam, The University of Hong Kong, China

Received: July 14, 2013; **Accepted:** October 11, 2013; **Published:** November 25, 2013

Copyright: © 2013 Abubakar et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This work was supported by the National Grand Program on Key Infectious Disease Control (2012ZX10001-006, 2013ZX10001-002), the China Ministry of Health, and the 973 National Key Basic Research Project (2012CB519005), Ministry of Science and Technology of P.R. China. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

* E-mail: jianqingxu2008@gmail.com

☉ These authors contributed equally to this work.

Introduction

Human immunodeficiency virus type 1 (HIV-1) sequences belonging to the pandemic group M are classified into nine subtypes (A–D, F–H, J, and K), six sub-subtypes (A1–A4, and F1–F2), and a variety of inter-subtype recombinant forms (Los Alamos HIV sequence database: <http://hiv-web.lanl.gov/>). CRF01_AE as one of HIV-1 circulating recombinant forms (CRF) is considered globally the fifth largest subtype and found in South and South-East Asia, East Asia and a small number in Central Africa [1]. Consequently, CRF01_AE had been reported as prevalent and responsible for the vast majority of infections in South and South East Asia [1,2]. In China, as part of East Asian region, CRF01_AE represents one of major HIV-1 genotypes reportedly associated with particular risk populations [1,3,4].

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 characteristics in these locations [5,6]. Moreover, CRF01_AE, which was the dominant strain in the sexual-risk population [6], apparently played more important role in MSM (men having sex with men) population [3]. Additionally, the increasing percentages of CRF01_AE infections in many Chinese cities exceeded those of the previously predominant subtype B [3,7,8].

So far, little is known about the phylogenetic relationships and epidemic history of CRF01_AE clades in Asia at large and China in particular. Furthermore, most studies in China had been conducted at provincial level [3,9–11]; therefore, much is needed to explore CRF01_AE epidemic history by involving all

Table 1. Region, country, number and sampling date of all sequences used in the study.

Region	Country (n)	N	Sampling date
Central Africa (CF)	Central African Rep. (CAR)	3	1990
	Democratic Republic of Congo (CD)	9	1984
West Africa (WA)	Cameroon (CM)	4	2007
South and South-East Asia (SSE)	Afghanistan (AF)	1	2007
East Asia (EA)	Indonesia (ID)	8	1993-2007
	Iran (IR)	1	2010
	Myanmar (MM)	10	1997-1999
	Singapore (SG)	28	1996
	Thailand (TH)	1474	1990-2009
East Asia (EA)	Vietnam (VN)	218	1997-2008
	China (CN)	182	1996-2010
East Asia (EA)	Japan (JP)	9	1993-2000
	West and Central Europe (WCE)	Denmark (DK)	2
West and Central Europe (WCE)	UK (GB)	1	2005
	Sweden (SE)	7	2006-2010
Total	15	1,957	1990-2010

doi: 10.1371/journal.pone.0080487.t001

Asian clades to obtain more comprehensive picture of CRF01_AE dynamics.

By using a combination of phylogenetic analyses and Bayesian coalescent-based approach in this study, we analyzed a total of 1,957 HIV-1 CRF01_AE *gag* p17 sequences from 5 global regions containing 15 countries including China (Table 1), to better understand the phylogenetic relationship between global and Asian clades, and to explore in greater detail the epidemic history of CRF01_AE in China.

Our results demonstrated the close phylogenetic relationship between Thailand, Vietnam and Chinese CRF01_AE sequences. Interestingly, we found multiple independent introductions and multiple independent autochthonous clades of HIV-1 CRF01_AE circulating in different geographic locations of China, which probably descended from Thailand, Vietnam or both. Yunnan province of China was identified as the earliest introduction site, and together with Guangxi, Fujian and Liaoning were demonstrated as locations harboring those clades. Those clades, however, were indicated as independently major or sporadic introductions, which likely took place from 1991 through 2003.7.

Materials and Methods

Sequence dataset and Alignment

All HIV-1 CRF01_AE *gag* p17 sequences matching the selected genomic region (nt 913–1111 relative to HXB2 clone) were retrieved from Los Alamos HIV Database (<http://hiv.lanl.gov>). Of note, HIV-1 *gag* p17 gene had been used previously for HIV-1 genotyping or phylogenetic analysis and appeared less prone to recombination [2,6,12-14]. Additionally,

sequences of this HIV-1 genomic region cover uniquely most of CRF01_AE prevalent Asian countries and key provinces in China as well (<http://hiv.lanl.gov>); a feature qualifies this genomic region to fit the study of CRF01_AE epidemic history in China. And, as a preliminary step to test the high-quality and eligibility of this genomic region, all *gag* p17 sequences were subjected to bioinformatic tools that available at Los Alamos database website.

Countries were regrouped into geographic regions according to the classification proposed by Hemelaar et al [1]. To improve the accuracy of phylogenetic inference, subtype assignment of all sequences was confirmed by the REGA HIV subtyping tool v.2[15]. All sequences with evidence of erroneous subtype assignment or evidence of inter-subtype recombination were excluded. The final dataset was composed of 1,957 CRF01_AE p17 sequences sampled from blood and collected between 1990-2010, which represented 15 countries (Table 1). Sequences were obtained through the utilization of PBMCs, serum or plasma samples of HIV-1 positive individuals, and methods of extraction and amplification were deliberately detailed in their relative studies [6,8,16-20]. All sequences were aligned with HIV-1 subtype reference sequences (<http://www.hiv.lanl.gov> sequence alignment page) using GeneCutter (<http://www.hiv.lanl.gov> GeneCutter page). Alignment quality was checked manually in Bio-Edit V7.8 [21] to ensure that the alignments did not contain obvious problems and that they were correctly codon aligned.

Phylogenetic analysis of HIV-1 CRF01_AE sequences

Maximum-likelihood (ML) phylogenetic tree was inferred under the GTR+I+C4 nucleotide substitution model (default model) after reconstruction with PhyML program [22] using an online web server [22]. Heuristic tree search was performed using the SPR branch-swapping algorithm and the reliability of the obtained topology was estimated with the approximate likelihood-ratio test (aLRT) [23] based on the Shimodaira-Hasegawa-like (SH-like) procedure. The ML trees were visualized using the FigTree v1.3.1 program [24]. Only lineages in ML tree supported by approximate likelihood ratio test (aLRT) ≥ 0.95 were selected. We assessed the temporal signal of p17 dataset as previously described in ML tree construction and clock-like behavior of the resultant data was also evaluated through regressing root-to-tip distance by using Path-O-Gen v1.3[25,26]. This analysis effectively provides a measure of the amount of variation in genetic distances explained by sampling time.

Phylogeographic Analysis of HIV-1 CRF01_AE lineages

The evolutionary rate (μ , units are nucleotide substitutions per site per year, subst./site/year), the age of the most recent common ancestor (tMRCA), and the probable geographic origin of Chinese CRF01_AE lineages were estimated using the Bayesian Markov Chain Monte Carlo (MCMC) approach implemented in BEAST software package v1.6.2 [27,28]. The best nucleotide substitution model for all datasets was evaluated by Mega software v5.2. MCMC analyses were performed using the Hasegawa-Kishino-Yano (HKY) nucleotide substitution model [29,30] with a gamma-distributed model

among site rate variation using four rate categories (C4) [29,31], an uncorrelated Lognormal relaxed molecular clock model [32], and Bayesian Skyline coalescent tree prior [33]. Two separate MCMC chains were run for 8×10^8 generations and adequate chain mixing was checked by calculating the effective sample size (ESS) after excluding an initial 10% for each run using program TRACER v1.4 [34]. MCMC runs converged to almost identical values and combined estimates showed ESS values >150 . Maximum clade credibility (MCC) trees were summarized from the posterior distribution of trees with TreeAnnotator and visualized with FigTree v1.3.1. For the exclusion that substitution rate and tMRCA estimates could be biased by either over- or under-representation of countries under study, BEAST analysis was applied to a dataset of randomly selected p17 sequences containing equal number of sequences (i.e. 15 each) per country in case of countries with abundant sequences, plus 3 sequences from less represented countries. The randomization process was repeated several times and nucleotide substitution rates and tMRCA were estimated every time [26]. Furthermore, to confirm the latter outcome, the full gag (HXB2 numbering: 790-2235) dataset of the same randomly selected dataset was again compared with both gag p17 datasets. Once again no significant difference between parameters was detected (Data not shown).

Results

Sequences collected from database and used for the analyses

Among all HIV-1 CRF01_AE *gag* p17 gene sequences that were downloaded, we first discarded all sequences without geographical information or identified as redundant sequences. Sequences which were not identified as CRF01_AE (i.e. intersubtype recombinants) by the REGA-subtyping tool were removed as well; and we kept only one sequence per patient. Finally, the utilized dataset included a total of 1,957 CRF01_AE *gag* p17 (nt 913–1111 relative to HXB2 clone), covering all five geographic regions included in the study (West Africa, Central Africa, West and Central Europe, East Asia, and South and Southeast Asia) (Table 1). This dataset was employed to characterize the phylogenetic relationships between CRF01_AE sequences of all involved regions, and to delineate with more confidence the relationships within Asian sequences, which was the main focus of the study.

Phylogenetic relationships between HIV-1 CRF01_AE sequences

Maximum likelihood tree was reconstructed using a total of 1,957 CRF01_AE *gag* p17 sequences from all 5 global regions (Figure 1). Figure 1 demonstrates the maximum likelihood (PhyML) tree, indicating that all CRF01_AE strains were originally from Central Africa and sequences from South and Southeast Asia formed the earliest Asian subcluster. Of note, all sequences of the latter subcluster were Thai sequences. Thus, Thailand represented the origin of all Asian CRF01_AE strains. The highlighted lineages in grey represent 27 lineages (L1-L27) that highly supported by aLRT value ≥ 0.95 at the first or second node [35]. Interestingly, all lineages ($n=27$)

contained interspersed Asian sequences mainly from South and Southeast Asia (SSE) region and from East Asia, indicating close phylogenetic relationships between these two regions. However, only one lineage (i.e. L7) contained in addition to the above-mentioned regions one sequence from West and Central Europe region (i.e. GB). Of more interest, 13 lineages composed only of one Chinese sequence, while the remaining 14 lineages contained ≥ 2 Chinese sequences (designated major Chinese lineages).

A closer view of those 14 Asian lineages showed that Chinese sequences branched with sequences from Thailand, Vietnam or both, demonstrating close phylogenetic relationships between those CRF01_AE sequences. Our findings are in agreement with previous studies indicating HIV-1 CRF01_AE transmission involving the 3 countries [6,17]. Moreover, major Chinese clades phylogenetically related to Thai sequences found in lineages L16, L17, L22, L23 and L27, while those related to Vietnam contained in lineages 6 and 20 (Figure S1).

Furthermore, those Asian lineages contained 5 major Chinese subclusters in which the latter branched with Thai sequences and represented sequences from Guangxi (southern Chinese province bordering Vietnam) (L16, L17), Hubei (central province) (L22) and 2 clades from Liaoning (northeastern Chinese province) (L27). Whereas, another 2 Chinese subclusters branched with Vietnam sequences represented Fujian (southeastern province) (L6) and Guangxi (L20) sequences.

Our results pointed directly to close phylogenetic relationships between CRF01_AE clades from Thailand or Vietnam, and sequences of the above mentioned geographic locations inside China (Figure S1), an outcome that is consistent with previous epidemiologic reports [36,37].

Origin of HIV-1 CRF01_AE sequences circulating in China

To determine with more confidence the most probable origin of those Chinese CRF01_AE lineages, we applied a Bayesian statistical framework that allows ancestral reconstruction of the locations at the interior nodes of Bayesian tree while accommodating phylogenetic uncertainty. We combined Chinese CRF01_AE sequences and those from countries previously pointed out by ML tree as being the most probable ancestral locations of Chinese lineages (Central African Republic, Thailand and Vietnam) to perform this analysis. For Chinese sequences, we included only sequences of China's peripheral provinces, wherein earliest CRF01_AE infections in China were identified and probably constituted the entry sites. The selected geographic locations included Yunnan ($n=6$), Guangxi ($n=47$), Fujian ($n=8$) and Liaoning ($n=47$). Global sequences were those that branched with one or more Chinese sequences until the second ancestral node in the ML phylogenetic tree, which included Thailand ($n=147$), Vietnam ($n=20$) and CAR ($n=3$). Sequences from other countries were excluded due to poor representation (≤ 2 sequences). Ultimately, a total of 278 sequences were used for the reconstruction of Bayesian maximum clade credibility (MCC) tree.

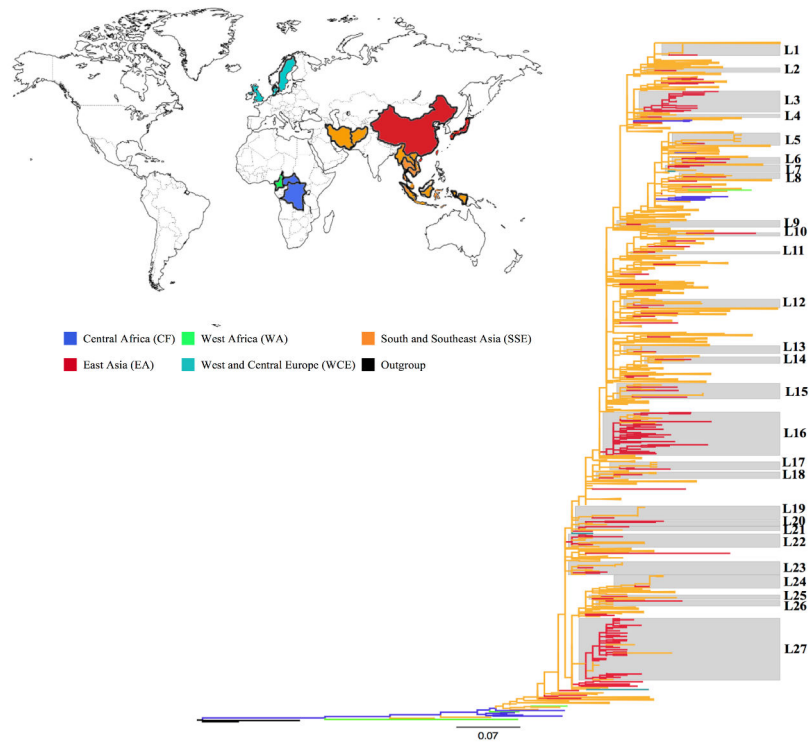


Figure 1. ML tree of HIV-1 CRF01_AE gag p17 sequences from 5 global regions (15 countries). The boxes highlight the position of the Chinese CRF01_AE lineages. The color of the branches represents the geographic region from where the CRF01_AE sequences originated, according to the map given at the top left of the figure. For visual clarity, some clades comprised of sequences sampled from SSE region were collapsed in orange triangles. The tree was rooted using outgroup (black branches). Horizontal branch lengths are drawn to scale with the bar at the bottom indicating nucleotide substitutions per site.

doi: 10.1371/journal.pone.0080487.g001

Figure 2A summarizes the overall topology of the obtained Bayesian MCC tree. It included 13 Asian lineages (L1-L13) in grey color that contained Chinese sequences. Figure 2B displays a closer view of the 13 lineages; interestingly, 10 major Chinese clades (≥ 2 sequences) contained within 8 Asian lineages represented most likely independent CRF01_AE clades.

Lineage 1 contained 2 independent Chinese clades from Liaoning, which branched with Thai sequences. However, lineage 1 is poorly supported (Posterior Probability (PP) value=0.54), raising uncertainty about those 2 clades as descendants from Thailand. Another Liaoning subcluster formed within lineage 11, though branched with Thai sequences, the node was not supported (PP<0.50).

Lineages 2 and 4 composed of independent Chinese clades from Guangxi, which descended from Vietnam with high support at the first or second node i.e. PP = 1 and 0.85, respectively. Furthermore, another 2 Guangxi clades descended from Thailand demonstrated in lineages 5 and 6 with high support (PP value = 0.90 and 0.99, respectively). Together, our results pointed to 4 major Chinese CRF01_AE clades from Guangxi most likely introduced from Thailand (n=2) and Vietnam (n=2). Other major Guangxi clades observed in lineages 4 and 9, but without PP support (PP=0.30, 0.32,

respectively), which likely resulted from internal spread network.

In addition, lineage 13 contained another independent Chinese clade from Fujian that descended from Vietnam, and supported by a PP value equal to 0.81 at the second node.

Altogether, ancestral reconstruction of the locations at the interior nodes of Bayesian tree revealed that the most probable place at the root of those major Chinese CRF01_AE clades could be traced to a neighboring country with high confidence including: Thailand for Guangxi clades in lineages 5 and 6; and Vietnam for Guangxi clades in lineages 2 and 4, and Fujian clade in lineage 13. The phylogenetic relationships between the 4 mentioned provinces/regions of China, Thailand and Vietnam were documented previously [3,6,37,38].

Time-scale of CRF01_AE introductions into China

By employing the time-scale Bayesian MCC tree, we estimated the age of the MRCA (*tMRCA*) of those Chinese CRF01_AE clades in all 13 Asian lineages. As mentioned above, we considered Chinese clades with ≥ 2 sequences in lineages as major introductions and <2 as sporadic introductions into China. The coefficient of substitution rate variation was higher than zero (0.56; 95% HPD: 0.32–0.79),

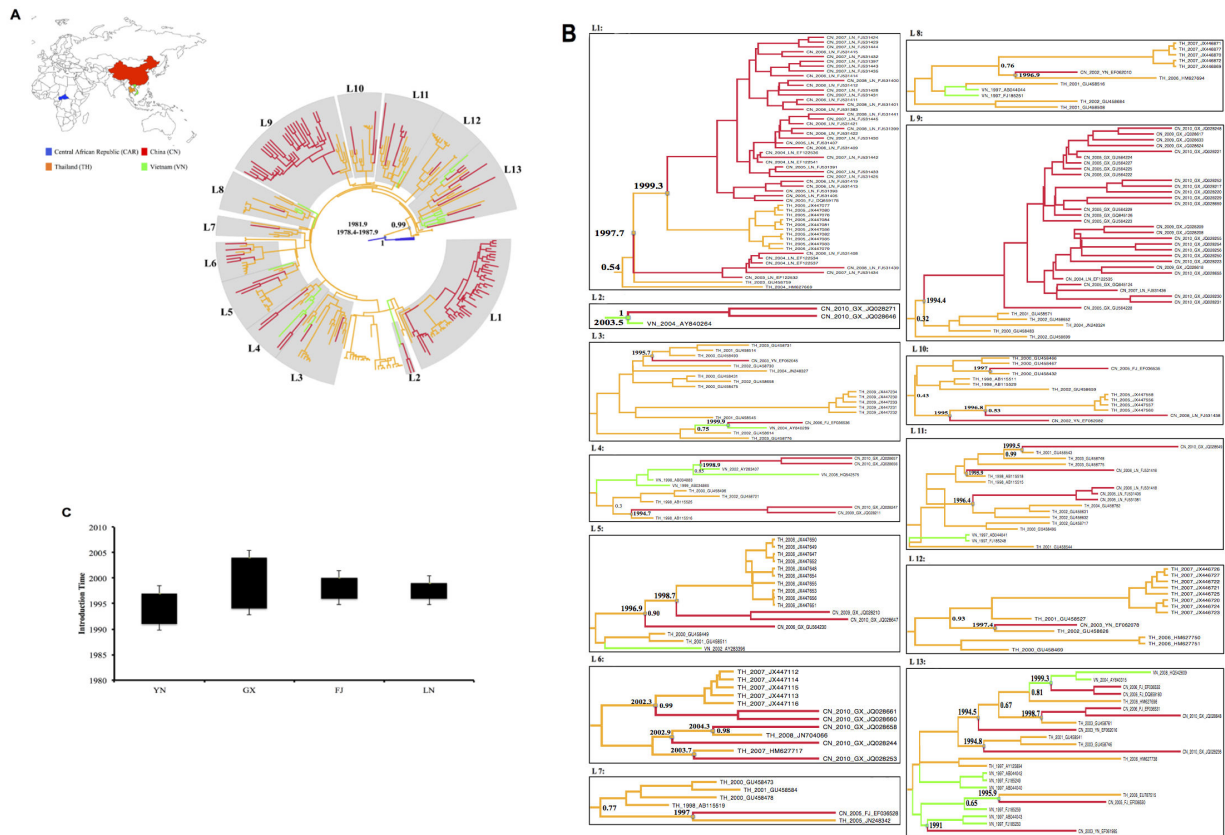


Figure 2. Time-scaled Bayesian Maximum Clade Credibility (MCC) tree. **A:** Overall topology of MCC tree of the Chinese CRF01_AE lineages and the most closely related global sequences. Branches are colored according to the most probable location state of their descendent nodes. The color of the branches represents the geographic region from where the CRF01_AE sequences originated, according to the map given at the top left of the figure. The boxes highlight the position of the Chinese CRF01_AE clades. The first (CF) and second (TH) nodes representing initial nodes lineages to all lineages together with posterior probability values 1 and 0.99, respectively, are indicated. The median and 95% HPD interval of the tMRC at the root of the tree is indicated. The tree was automatically rooted under the assumption of a relaxed molecular clock.

B: Close view of the Chinese CRF01_AE clades and the most closely related Asian sequences in MCC tree. The color of branches represents the geographic region from where the sequence originated. The names of CRF01_AE strains include reference to country origin, year of isolation, province (Chinese provinces) and GenBank Accession number. Countries represented are Central African Republic (CF), Thailand (TH), Vietnam (VN) and China (CN), the later represented by provincial sequences: Yunnan (YN), Guangxi (GX), Fujian (FJ) and Liaoning (LN). The posterior probability (PP) support values are indicated only at key nodes.

C: Estimated time-range of HIV-1 CRF01_AE earliest introductions into 4 border provinces (Yunnan (YN), Guangxi (GX), Fujian (FJ) and Liaoning (LN)) in China, the 95% highest posterior density (HPD) credible regions are provided as the range of state tMRC.

doi: 10.1371/journal.pone.0080487.g002

which supported the use of a relaxed molecular clock model. The rate of evolution of gag p17 gene was estimated to be 2.58×10^{-3} substitutions per site per year.

The tMRC of the Asian CRF01_AE lineage was estimated at 1981.9 (95%, CI: 1978.4-1987.9), which is within the range of previous estimations [3].

The tMRCs of major Chinese clades were estimated at: 1996.4 (95%, CI: 1994.5-1999.7) and 1997.7 (95%, CI: 1996.1-2000.7), and 1999.3 (95%, CI: 1996.2-2000) for Liaoning clades in lineage 11 and 1, respectively; 1994.4 (95%, CI: 1995.3-1997.4), 1994.7 (95%, CI: 1993.3-1997.1), 1998.7

(95%, CI: 1997.4-2004), 1998.9 (95%, CI: 1997.8-2001.6), 2002.3 (95%, CI: 2001.2-2006.1) and 2003.5 (95%, CI: 2001.5-2006) for Guangxi clades in lineages 9, 4, 5, 4, 6 and 2, respectively; 1999.3 (95%, CI: 1998.8-2003.3) for Fujian in lineage 13 (Figure 2B). The estimated introduction time of those major lineages is consistent with previous epidemiological data and reports [3,11,37,39].

Figure 2C summarizes the time-range of probable CRF01_AE introductions into China including all clades together (major and sporadic). Interestingly, a Chinese clade from Yunnan (southwestern province bordering Myanmar, Laos

and Vietnam) in lineage 13 emerged as the earliest CRF01_AE clade in China and was dated back to 1991 (95%, CI: 1992.4-1994.2). Although this clade (one sequence) was not supported in lineage 13, the tMRCA estimation coincided with recently published study (i.e. 1990.9)[40]; even though, it should be considered with caution.

Discussion

In this study we analyzed HIV-1 CRF01_AE *gag p17* sequences from 5 global regions (15 countries) to better understand the phylogenetic relationships and epidemic history of this CRF in Asia and China in particular. The analysis of 1,957 globally distributed sequences most likely provided an adequate representation of global CRF01_AE diversity, and provided not only additional information on the subtype epidemic in Asia, but also in other parts of the world. Our study revealed that all CRF01_AE strains were originated in Central African Republic and then spread into South and Southeast Asian region, wherein existed an important triangle of transmission route i.e. Thailand - Vietnam - China, and Thailand - China, leading obviously to multiple independent introductions of CRF01_AE into China from neighboring countries.

The earliest introduction probably occurred in 1991 (Lineage 13) as a sporadic event from Vietnam to Yunnan province, which seemingly not the cause of an outbreak or spread to other regions, even though this date should be considered cautiously. Guangxi represented the second earliest site to receive HIV-1 CRF01_AE in 1994.4, which resulted in an outbreak in Guangxi and formed the major cluster (Lineage 9) in MCC tree; Interestingly, it remains uncertain where the sequences were derived from since the PP value for Thailand-Guangxi subcluster was 0.32. Subsequently, Guangxi received further independent transmissions from both Vietnam in 1998.9 and 2003.5, and Thailand in 1994.7, 1998.7 and 2002.3, indicating frequent travelling between Guangxi and Thailand or Vietnam. Liaoning apparently was the third site of CRF01_AE introduction in which major introductions took place in 1996.4, 1997.7 and 1999.3. Again, the origins of those independent introductions remain uncertain because the PP value for Thailand - Liaoning was below 0.5. Perhaps, Liaoning clades in lineages 1 and 9 were derived from earliest strains of Yunnan; however, the linkages were not significantly established due to the scarcity of sequences from Yunnan (only 6 sequences were used) at the time of the study. Fujian represented the fourth site to receive CRF01_AE introduction in 1999.3 from Vietnam. These observations are in accordance with previous reports [3,6,37,41-43]. It is worth mentioning that lack of sequences from unrepresented countries bordering China, in addition to over-representation of others probably impeded the study to illustrate more phylogenetic relationships and introductions.

As widely known, the spread of HIV-1 variants is mainly driven by means of sexual contact including both heterosexual [44] and homosexual [3], in addition to blood donation, injection drug use and mother to child transmission [45], which are known routes of transmission in Asia. CRF01_AE transmission

in particular has been related remarkably to sexual contact [16,45,46]. Hence, as commercial sex is considered to be an occasional practice, it had most likely contributed to the earliest sporadic transmissions and interspersed linkages between CRF01_AE variants, an observation that indicated by our study and previous reports [6,16,41,43,46]. Nevertheless, the current study was precluded from deriving a route-linked conclusion on lineages, due to the lack or limitation of information describing patients' transmission routes in HIV database. However, as the travelling population could significantly influence the transmission patterns, multiple introductions of CRF01_AE from Vietnam and Thailand were likely driven by the increasing inter-country travels.

It remains unclear why only two large clusters were formed in spite of many independent introductions. The established transmission from neighboring countries into China with high support failed to form any large cluster (>2 sequences), indicating the unlikeliness of those strains to cause any significant outbreak. In contrast, we were unable to directly link the two large clusters of Guangxi and Liaoning to any strains in neighboring countries, suggesting that those two clusters were most likely autochthonous clades of uncertain origin, knowingly a prolonged evolution is required for those strains to exist as fully imprinted Chinese strains. In this regard, it is critical to identify the populous strains circulating among the population for the development of effective vaccine and microbicides. And certainly further surveillances to explore the biological differences between strains derived from large clusters and sporadically transmitted strains are highly needed.

In conclusion, this study demonstrates the existence of multiple independent introductions and several autochthonous transmission networks of HIV-1 CRF01_AE in China that probably resulted in more complicated pattern of CRF01_AE transmission in this most populous nation. Our analyses not only deepen the understanding on the epidemic history of CRF01_AE in Asia, but also drive the attention to the paramount importance of continuous surveillance of CRF01_AE variants in China to adapt treatment, vaccine and microbicide strategies.

Supporting Information

Figure S1. Close view of ML lineages containing Chinese CRF01_AE clades and the most closely related global sequences. The color of branches represents the geographic region from where the sequence originated, as explained in Figure 1. The names of CRF01_AE strains include reference to geographic region, country of origin, year of isolation, province (in case of Chinese sequences) and GenBank Accession number. Geographic regions represented are South and Southeast Asia (SSE), West and Central Europe (WCE), and East Asia (EA). Countries represented are Thailand (TH), Vietnam (VN), Japan (JP), Singapore (SG), Myanmar (MM), Great Britain (GB), and China (CN). Chinese provinces represented are Yunnan (YN), Guangxi (GX), Fujian (FJ), Liaoning (LN), Hebei (HEB), Hubei (HUB) and Henan (HEN). The aLRT support values are indicated only at key nodes. (DOCX)

Author Contributions

Performed the experiments: YA ZM. Analyzed the data: YA. Wrote the manuscript: YA. Revised the manuscript: JX. Conceived the study: JX. Designed the experiments: JX XZ.

References

- Hemelaar J, Gouws E, Ghys PD, Osmanov S (2011) Global trends in molecular epidemiology of HIV-1 during 2000–2007. *AIDS* 25: 679–689. doi:10.1097/QAD.0b013e328342ff93. PubMed: 21297424.
- Han Z, Leung TW, Zhao J, Wang M, Fan L et al. (2009) A HIV-1 heterosexual transmission chain in Guangzhou, China: a molecular epidemiological study. *Virology* 6: 148. doi:10.1186/1743-422X-6-148. PubMed: 19778458.
- An M, Han X, Xu J, Chu Z, Jia M et al. (2012) Reconstituting the Epidemic History of HIV Strain CRF01_AE among Men Who Have Sex with Men (MSM) in Liaoning, Northeastern China: Implications for the Expanding Epidemic among MSM in China. *J Virol* 86: 12402–12406. doi:10.1128/JVI.00262-12. PubMed: 22933290.
- Shao Y (2001) HIV/AIDS: perspective on China. *AIDS Patient Care STDS* 15: 431–432. doi:10.1089/108729101316914449. PubMed: 11522218.
- Zeng H, Sun Z, Liang S, Li L, Jiang Y et al. (2011) Emergence of a New HIV-1 CRF01_AE Variant in Guangxi, Southern China. *AIDS Research and Human Retroviruses* 10.
- Zhang Y, Lu L, Ba 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. doi: 10.1371/journal.pmed.0030443. PubMed: 17105339.
- Han X, Dai D, Zhao B, Liu J, Ding H et al. (2010) Genetic and epidemiologic characterization of HIV-1 infection in Liaoning Province, China. *J - Journal of Acquir Immune Defic Syndr* 53 Suppl 1: S27–S33. doi:10.1097/QAI.0b013e3181c7d5bf.
- Li L, Lu 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. doi:10.1016/j.meegid.2011.05.017. PubMed: 21645646.
- Zeng H, Sun Z, Liang S, Li L, Jiang Y et al. (2012) Emergence of a new HIV type 1 CRF01_AE variant in Guangxi, Southern China. *AIDS Res Hum Retroviruses* 28: 1352–1356. doi:10.1089/aid.2011.0364. PubMed: 22264007.
- Guo D, Ding N, Xu Y, Guo H, Wei JF et al. (2009) 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* 25: 619–623. doi:10.1089/aid.2009.0002. PubMed: 19534629.
- Chen JH, Wong KH, Li P, Chan KC, Lee MP et al. (2009) Molecular epidemiological study of HIV-1 CRF01_AE transmission in Hong Kong. *J Acquir Immune Defic Syndr* 51: 530–535. doi:10.1097/QAI.0b013e3181aac516. PubMed: 19521252.
- Pang W, Zhang C, Duo L, Zhou YH, Yao ZH, et al. (2012) Extensive and complex HIV-1 recombination between B', C and CRF01_AE among IDUs in Northern Myanmar. *AIDS* 000–000
- Sánchez-Merino V, Herrero CC, Amorin-Nink A, von Briesen H, López-Galíndez C (2003) Genetic analysis of culture-negative UNAIDS subtype C samples. *AIDS Res Hum Retroviruses* 19: 49–55. doi: 10.1089/08892220360473961. PubMed: 12596720.
- Zhang L, Chen Z, Cao Y, Yu J, Li G et al. (2004) Molecular characterization of human immunodeficiency virus type 1 and hepatitis C virus in paid blood donors and injection drug users in China. *J Virol* 78: 13591–13599. doi:10.1128/JVI.78.24.13591-13599.2004. PubMed: 15564470.
- De Oliveira T, Deforche K, Cassol S, Salminen M, Paraskevis D et al. (2005) An automated genotyping system for analysis of HIV-1 and other microbial sequences. *Bioinformatics* 21: 3797–3800. doi:10.1093/bioinformatics/bti607. PubMed: 16076886.
- Li L, Sun G, Li T, Liu Y, Chen L et al. (2012) Multiple introductions of HIV into men who have sex with men were found in Zhengzhou City, China. *AIDS Res Hum Retroviruses* 28: 1147–1151. PubMed: 22236149.
- 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. doi:10.1089/aid.2010.0026. PubMed: 20528151.
- Gesprasert G, Wichukchinda N, Mori M, Shiino T, Auwanit W et al. (2010) HLA-associated immune pressure on Gag protein in CRF01_AE-infected individuals and its association with plasma viral load. *PLOS ONE* 5: e11179. doi:10.1371/journal.pone.0011179. PubMed: 20567513.
- Jinnopat P, Isarangkura-na-ayuthaya P, Utachee P, Kitagawa Y, de Silva UC, et al. (2009) Impact of amino acid variations in Gag and protease of HIV type 1 CRF01_AE strains on drug susceptibility of virus to protease inhibitors. *J Acquir Immune Defic Syndr* 52: 320–328
- Tran TH, Maljkovic I, Swartling S, Phung DC, Chiodi F et al. (2004) HIV-1 CRF01_AE in intravenous drug users in Hanoi, Vietnam. *AIDS Res Hum Retroviruses* 20: 341–345. doi: 10.1089/088922204322996581. PubMed: 15117458.
- Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp Ser* 41: 95–98.
- Guindon S, Gascuel O (2003) A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol* 52: 696–704. doi:10.1080/10635150390235520. PubMed: 14530136.
- Anisimova M, Gascuel O (2006) Approximate likelihood-ratio test for branches: A fast, accurate, and powerful alternative. *Syst Biol* 55: 539–552. doi:10.1080/10635150600755453. PubMed: 16785212.
- Rambaut A (2009) FigTree v1.3.1: Tree Figure Drawing Tool. Available <http://treebioedacuk/software/figtree/>. Accessed 2012 April 20
- Rambaut A (2009) Path-O-Gen: temporal signal investigation tool. Available online at: <http://tree.bio.ed.ac.uk/software/pathogen/>. Accessed 2013 June 13
- Pagán I, Holguín A (2013) Reconstructing the Timing and Dispersion Routes of HIV-1 Subtype B Epidemics in The Caribbean and Central America: A Phylogenetic Story. *PLOS ONE* 8: e69218. doi:10.1371/journal.pone.0069218. PubMed: 23874917.
- Drummond AJ, Nicholls GK, Rodrigo AG, Solomon W (2002) Estimating mutation parameters, population history and genealogy simultaneously from temporally spaced sequence data. *Genetics* 161: 1307–1320. PubMed: 12136032.
- Drummond AJ, Rambaut A (2007) BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol Biol* 7: 214. doi: 10.1186/1471-2148-7-214. PubMed: 17996036.
- Meng Z, Xin R, Zhong P, Zhang C, Abubakar YF et al. (2012) A New Migration Map of HIV-1 CRF07_BC in China: Analysis of Sequences from 12 Provinces over a Decade. *PLOS ONE* 7: e52373. doi:10.1371/journal.pone.0052373. PubMed: 23300654.
- Hasegawa M, Kishino H, Yano T (1985) Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *J Mol Evol* 22: 160–174. doi:10.1007/BF02101694.
- Yang Z (1994) Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: approximate methods. *J Mol Evol* 39: 306–314. doi:10.1007/BF00160154. PubMed: 7932792.
- Drummond AJ, Ho SY, Phillips MJ, Rambaut A (2006) Relaxed phylogenetics and dating with confidence. *PLoS Biol* 4: e88. doi: 10.1371/journal.pbio.0040088. PubMed: 16683862.
- Drummond AJ, Rambaut A, Shapiro B, Pybus OG (2005) Bayesian coalescent inference of past population dynamics from molecular sequences. *Mol Biol Evol* 22: 1185–1192. doi:10.1093/molbev/msi103. PubMed: 15703244.
- Rambaut A, Drummond A (2007) Tracer v1.4. Available <http://beastbioedacuk/Tracer>. Accessed 2012 April 20
- Afonso JM, Morgado MG, Bello G (2012) Evidence of multiple introductions of HIV-1 subtype C in Angola. *Infect Genet Evol* 12: 1458–1465. doi:10.1016/j.meegid.2012.05.005. PubMed: 22634597.
- Han Z, Leung TW, Zhao J, Wang M, Fan L et al. (2009) A HIV-1 heterosexual transmission chain in Guangzhou, China: a molecular epidemiological study. *Virology* 6: 148. doi:10.1186/1743-422X-6-148. PubMed: 19778458.
- Feng Y, He X, Jenny H, Li F, Li X et al. (2013) The rapidly expanding CRF01_AE epidemic in China is driven by multiple lineages of HIV-1 viruses introduced in the 1990s. *AIDS* 27: 1793–1802. doi:10.1097/QAD.0b013e328360db2d. PubMed: 23807275.

38. Huang HL, Yan PP, Zheng J, Wu SL, Cheng G et al. (2006) Genetic characterization of three CRF01_AE full-length HIV type 1 sequences from Fujian Province, China. *Chin Med J (Engl)* 119: 1622-1628. PubMed: 17042975.
39. Han X, An M, Zhang M, Zhao B, Wu H et al. (2013) Identification of 3 Distinct HIV-1 Founding Strains Responsible for Expanding Epidemic Among Men Who Have Sex With Men in 9 Chinese Cities - *J Acquir Immune Defic Syndr* 64: 16-24. doi:10.1097/QAI.0b013e3182932210.
40. Ye J, Xin R, Yu S, Bai L, Wang W et al. (2013) Phylogenetic and Temporal Dynamics of Human Immunodeficiency Virus Type 1 CRF01_AE in China. *PLOS ONE* 8: e54238. doi:10.1371/journal.pone.0054238. PubMed: 23365653.
41. Chen M, Yang L, Ma YL, Su Y, Yang C et al. (2013) Emerging variability in HIV-1 genetics among recently infected individuals in Yunnan, China. *PLOS ONE* 8: e60101. doi:10.1371/journal.pone.0060101. PubMed: 23555898.
42. Kayoko K, Teiichiro S, Shigero K, Hironori S, Kyoko N et al. (1999) Genetic Similarity of HIV Type 1 Subtype E in a Recent Outbreak among Injecting Drug Users in Northern Vietnam to Strains in Guangxi Province of Southern China. *AIDS Res Hum Retroviruses*, 15: 1157-1168. PubMed: 10480629.
43. Li L, Chen L, Liang S, Liu W, Li T et al. (2013) Subtype CRF01_AE dominate the sexually transmitted human immunodeficiency virus type 1 epidemic in Guangxi, China. *J - Journal of Med Virol* 85: 388-395. doi: 10.1002/jmv.23360.
44. Gao F, Robertson DL, Morrison SG, Hui H, Craig S et al. (1996) The heterosexual human immunodeficiency virus type 1 epidemic in Thailand is caused by an intersubtype (A/E) recombinant of African origin. *J Virol* 70: 7013-7029. PubMed: 8794346.
45. Lu L, Jia MH, Zhang XB, Luo HB, Ma YL et al. (2004) Analysis for epidemic trend of acquired immunodeficiency syndrome in Yunnan Province of China. *Zhonghua Yu Fang Yi Xue Za Zhi* 38: 309-312. PubMed: 15498242.
46. Yu XF, Wang X, Mao P, Wang S, Li Z et al. (2003) Characterization of HIV type 1 heterosexual transmission in Yunnan, China. *AIDS Res Hum Retroviruses* 19: 1051-1055. doi:10.1089/088922203322588422. PubMed: 14686326.