OPEN

The rapidly expanding CRF01_AE epidemic in China is driven by multiple lineages of HIV-1 viruses introduced in the 1990s

Yi Feng[a,*], Xiang He[a,*], Jenny H. Hsi[a], Fan Li[a], Xingguang Li[a], Quan Wang[b], Yuhua Ruan[a], Hui Xing[a], Tommy Tsan-Yuk Lam[c], Oliver G. Pybus[c], Yutaka Takebe[a,d] and Yiming Shao[a]

Objectives: We sought to comprehensively analyze the origin, transmission patterns and sub-epidemic clusters of the HIV-1 CRF01_AE strains in China.

Methods: Available HIV-1 CRF01_AE samples indentified in national molecular epidemiologic surveys were used to generate near full-length genome (NFLG) sequences. The new and globally available CRF01_AE NFLG sequences were subjected to phylogenetic and Bayesian molecular clock analyses, and combined with epidemiologic data to elucidate the history of CRF01_AE transmission in China.

Results: We generated 75 new CRF01_AE NFLG sequences from various risk populations covering all major CRF01_AE epidemic regions in China. Seven distinct phylogenetic clusters of CRF01_AE were identified. Clusters 1, 2 and 3 were prevalent among heterosexuals and IDUs in southern and southwestern provinces. Clusters 4 and 5 were found primarily among MSM in major northern cities. Clusters 6 and 7 were only detected among heterosexuals in two southeast and southwest provinces. Molecular clock analysis indicated that all CRF01_AE clusters were introduced from Southeast Asia in the 1990s, coinciding with the peak of Thailand's HIV epidemic and the initiation of China's free overseas travel policy for their citizens, which started with Thailand as the first destination country.

Conclusion: China's HIV-1 epidemic of sexual transmissions, was initiated by multi-lineages of CRF01_AE strains, in contrast to the mono-lineage epidemic of B′ strain in former plasma donors and IDUs. Our study underscores the difficulty in controlling HIV-1 sexual transmission compared with parenteral transmission.

AIDS 2013, **27**:1793–1802

Keywords: China, CRF01_AE, HIV-1, near full-length genome, phylogenetic cluster, risk population

Introduction

The HIV-1 circulating recombinant form (CRF) 01_AE, first identified among female sex workers in northern Thailand in 1989 [1–3] initially named subtype (or clade) E before its identification as a new recombinant virus [4,5]. Phylogenetic studies have shown that CRF01_AE originated in central Africa in the 1970s and spread in

[a]State Key Laboratory for Infectious Disease Prevention and Control, National Center for AIDS/STD Control and Prevention, Chinese Center for Disease Control and Prevention, Collaborative Innovation Center for Diagnosis and Treatment of Infectious Diseases, Beijing, [b]TEDA School of Biological Sciences and Biotechnology, Nankai University, Tianjin, China, [c]Department of Zoology, University of Oxford, Oxford, UK, and [d]AIDS Research Center, National Institute of Infectious Diseases, Tokyo, Japan.

Correspondence to Dr Yiming Shao, Division of Research on Virology and Immunology, National Center for AIDS/STD Control and Prevention, China CDC, Beijing 102206, China.

Tel: +86 10 5890 0981; e-mail: yshao08@gmail.com

[*] Yi Feng and Xiang He contributed equally to the writing of this work.

Received: 27 November 2012; revised: 7 February 2013; accepted: 4 March 2013.

DOI:10.1097/QAD.0b013e328360db2d

ISSN 0269-9370

Thailand in the 1980s through heterosexual transmission [4,6]. It was subsequently confirmed as the first large-scale epidemic of a recombinant HIV-1 strain in the world [5,7].

In the early 1990s, CRF01_AE strains were first identified in China among persons at risk of sexual transmission and IDUs in the southwest provinces of Yunnan and Guangxi, which border Myanmar and Vietnam [7–11]. Subsequently, China's nationwide molecular epidemiological surveys in 1996–2002 showed that the early spread of CRF01_AE was limited to the eastern coastal areas and southwest border provinces, predominantly in heterosexual populations [12,13]. However, as of the latest nationwide molecular epidemiological survey for newly reported cases in 2006, CRF01_AE strains had quickly emerged as the most widespread HIV-1 strain in terms of geographic and risk group distributions, accounting for approximately 28% of nationwide HIV infections of that period [14]. More recently, CRF01_AE strains have also become the predominant strain among MSM in China [15,16]. Clearer delineation of the origin, timescale, spatial spread and risk population structure of the HIV-1 recombinant CRF01_AE epidemic in China is, therefore, of considerable public health importance.

In the present study, we sampled various risk populations across major CRF01_AE epidemic regions in China and determined a total of 75 new near full-length genome (NFLG) sequences of CRF01_AE strains. We analyzed the sequences using phylogenetic and molecular clock methods, and identified at least seven closely timed transmission clusters with distinct geographic and risk group distributions. Our findings elucidate the origins and complexity of China's CRF01_AE epidemic and underscore the challenges in controlling sexual transmission of HIV-1.

Methods

Sample selection and dataset characteristics

On the basis of the results of the latest national HIV molecular epidemiology survey (NHMES) [14], all HIV-1 CRF01_AE samples identified were subjects for this study. Approximately 50% of the available specimens ($n = 162$) were randomly selected for amplification of the NFLGs. Additional samples ($n = 99$) were collected in order to cover the regions where the NHMES study failed to obtain NFLG CRF01_AE sequences. All together, a total of 75 HIV-1 CRF01_AE NFLG sequences were obtained with about two-thirds ($n = 47$) from the NHMES and one-third ($n = 28$) from additional sampling. All available CRF01_AE NFLGs from China ($n = 27$) in the Los Alamos HIV database were also downloaded and analyzed together with the 75 newly generated NFLG sequences.

The dataset of 102 NFLGs of CRF01_AE strains consists of all major risk populations for CRF01_AE in China, including heterosexuals ($n = 58$), IDUs ($n = 20$), MSM ($n = 18$), and unknown risk ($n = 6$). Altogether, it provides wide coverage in geographical regions, from which over 90% of the total estimated CRF01_AE HIV-1-infected cases were found [14]. All participants, except four people (two from Yunnan province and two from Hunan provinces), in the study were antiviral treatment naive.

The study was approved by the institutional review board of the National Center for AIDS/STD Control and Prevention. Informed consent was obtained from all participants at the time of sample collection.

Near full-length genome amplification and sequencing

Viral RNA was extracted from plasma using the QIAamp Viral RNA Mini kit (Qiagen, Valencia, California, USA) and reverse transcribed by SuperScript III reverse transcriptase (Invitrogen, Carlsbad, California, USA). The NFLG fragments (552–9636 nt relative to HXB2) were amplified by nested PCR as described by Li *et al.* [17] using primers 01_AE-FL1.5 (5′- CCTTGAGTGC TTAAAGTGGTGTGTGCC CGTCTGT-3′, HXB2: 538–571) and 01_AE-FL1.3 (5′- ACCACTTTAAG CACTCAA GGCAAGCTTTATTG-3′, HXB2: 9666–9611) for the first round PCR reaction, and 01_AE-FL2.5 (5′- AGTGGTGTGTGCCCGTCTGTGTTAG GACTC -3′, HXB2: 552–581) and 01_AE-FL2.3 (5′- TTAAGCACTCAAGGCAAGCTTTATTGAGG CTT A-3′, HXB2: 9636–9604) for the second. The PCR products were purified using QIAquick Gel Extraction Kit (Qiagen, Valencia, California, USA) and direct-sequenced using BigDye terminators (Applied Biosystems, Foster City, California, USA). Sequences newly determined in this study are available in GenBank under accession number JX112796-JX112870.

Reference sequences and sequence alignment

In addition to the 102 Chinese NFLGs sequences described above, all available HIV-1 CRF01_AE NFLGs from other countries (accessed in April 2012) were downloaded from Los Alamos HIV sequence database as reference sequences for this analysis. After removing redundancy and those without sampling information, the references ($n = 92$) were combined with the Chinese sequences to yield a total dataset of 194 sequences. The geographic location, sampling year, and risk groups for the sequences used in this study are summarized in Table 1 and Supplemental Table S1, http://links.lww.com/QAD/A341. An initial alignment of all sequences was performed using Gene Cutter from the Los Alamos HIV sequence database (http://www.hiv.lanl.gov/content/sequence/GENE_CUTTER/cutter.html). The alignment was subsequently adjusted manually using BioEdit v7.0.9 [18].

Table 1. Geographic origin and risk group distribution of HIV-1 CRF01_AE strains used in the present study.[a]

Geographic source	Sampling year	*n*	Risk group[b]			
			Hetero	IDU	MSM	n/a
China		102 (27)	58 (21)	20 (4)	18	6
Beijing	2010	13			13	
Fujian	2005–2007	22 (12)	20 (11)	1 (1)		1
Guangdong	2007	10	4	5		1
Guangxi	1997–2007	25 (14)	17 (10)	8 (4)		
Guizhou	2007	6	2	2		2
Hunan	2010	2				2
Jiangsu	2007	5 (1)	4	1 (1)		
Jilin	2010	4	2		2	
Liaoning	2007	3	1		2	
Sichuan	2006	2	1	1		
Tianjin	2007	1			1	
Yunnan	2002–2009	9	7	2		
Afghanistan	2007	1 (1)				1 (1)
Central African Republic	1990	3 (3)				3 (3)
Hong Kong	2004	1 (1)				1 (1)
Indonesia	1993	1 (1)				1 (1)
Japan	1993	1 (1)	1 (1)			
Thailand	1993–2004	49 (49)				49 (49)
United States	1998–2000	3 (3)				3 (3)
Vietnam	1997–1998	33 (33)	17 (17)	16 (16)		
Grand Total		194 (119)	76 (39)	36 (22)	18	64 (58)

[a]The numbers of CRF01_AE NFLG sequences downloaded from the Los Alamos HIV sequence database are shown in parenthesis.
[b]Risk groups: Hetero, heterosexual; n/a, not available.

Model selection analyses

ModelTest v3.7 [19] was used to find the best-fitting model of nucleotide substitution for our dataset. The general time reversible (GTR) model with γ-distributed (Γ4) among-site rate heterogeneity, and a proportion of invariant sites (I) (the GTR+I+Γ4 model) were chosen as the most appropriate mode on the basis of the standard Akaike information criterion.

Phylogenetic analyses

PhyML 3.0 [20] was used to estimate a maximum likelihood phylogenetic tree for NFLG sequences using the GTR+I+Γ4 nucleotide substitution model. Tree topologies were searched heuristically using the subtree pruning and regrafting procedure. The confidence of each node in phylogenetic trees was determined using the bootstrap method with 500 replicates. The final maximum likelihood tree was visualized using the program FigTree v1.3.1 (http://beast.bio.ed.ac.uk).

Web-based software Hypermut (http://www.hiv.lanl.gov/) was used to screen for hypermutation from the 194 NFLGs. Consequently, one sequence (FJ061, Fig S1) was identified from Cluster 6 and removed to avoid disproportional influence on the Bayesian phylogenetic analysis.

To estimate the evolutionary rate (μ: nucleotide substitutions/site per year) and the divergence times (time to most recent common ancestor; tMRCA) of various CRF01_AE lineages, we performed Bayesian phylogeny analysis for the *gag*, *pol* and *env* genes. Phylogenies were inferred using BEAST v.1.6.1 under an uncorrelated log-normal relaxed clock model, GTR+Γ4 substitution model, and Bayesian skyline plot demographic model [21–23]. For each gene, BEAST analysis was performed using Markov Chain Monte Carlo runs of chain length 3×10^8, and the first 10–30% of states of each run were discarded as burn-in. Ten thousand trees were then sampled to estimate μ and the tMRCA for each CRF01_AE lineage. Convergence was checked using Tracer v1.5 (http://beast.bio.ed.ac.uk/Tracer), and most parameters had effective sample sizes more than 200, except the tMRCA of the *env* region of cluster 6 (ESS = 196) (Supplemental Table S2, http://links.lww.com/QAD/A341).

Results

Identification of seven independent HIV-1 CRF01_AE lineages in China

We generated 75 new NFLG sequences of CRF01_AE HIV-1 strains collected from various risk groups and geographical regions in China. We performed phylogenetic analysis on this dataset in combination with existing NFLG sequences available from the Los Alamos HIV Databases, which consisted of 27 sequences from China and 92 from other countries. As shown in Fig. 1, the maximum-likelihood phylogeny identified seven well supported (bootstrap values were greater than 90%) and distinct clusters of CRF01_AE strains in China. Detailed information of the phylogeny tree in Fig. 1 is shown in

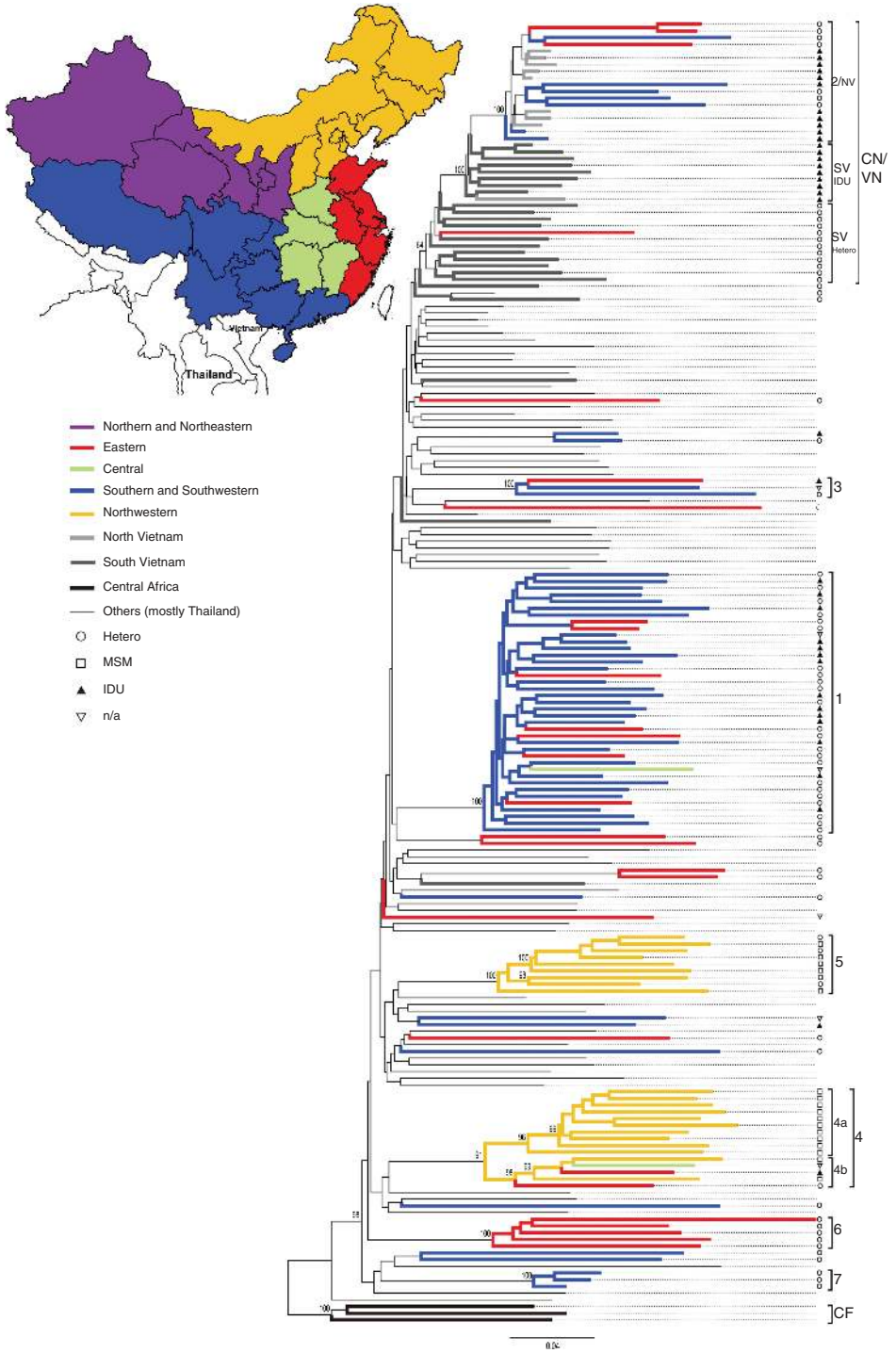

Fig. 1. HIV-1 CRF01_AE strains from China form seven distinct phylogenetic clusters. Near full-length genome (NFLG) sequences from China ($n = 102$; 75 were newly determined in this study) and other countries ($n = 92$) are analyzed in a maximum likelihood phylogeny. Sequences from China form seven unique phylogenetic clusters, labeled 1 through 7 in square brackets. Bootstrap values of all relevant major nodes are indicated. The geographic distribution of the strains is color-coded as shown in the inset. In addition, the CRF01_AE clusters of heterosexuals and IDUs in southern Vietnam (labeled with SV) are labeled to show their relationship to cluster 2; cluster 4 is subdivided into clusters 4a and 4b. A total of 18 CRF01_AE strains from southern provinces (categorized as 'ungrouped'; see Table 2) are in a Thailand CRF01_AE radiation.

supplemental Figure S1, http://links.lww.com/QAD/A341. Clusters 1 ($n = 39$), 2 ($n = 10$) and 3 ($n = 3$) were found primarily among heterosexuals and IDUs in southern provinces of China (Table 2). Of note, cluster 2 contained CRF01_AE strains circulating among IDUs in northern Vietnam ($n = 8$), southeastern China

Table 2. Classification and geographic distribution of distinct CRF01_AE phylogenetic clusters in China.[a]

CRF01_AE lineages	Province[b]	*n*	Risk group[c]				Remark
			Hetero	IDU	MSM	n/a	
Hetero/IDU clusters							
Cluster 1		39 (9)	23 (7)	14 (2)		2	
	Jiangsu	2	2				
	Sichuan	2	1	1			
	Fujian	5	5				
	Hunan	1				1	
	Guizhou	4	1	2		1	
	Guangxi	18 (9)	12 (7)	6 (2)			
	Guangdong	7	2	5			
Cluster 2		10 (5)	8 (3)	2 (2)			
	Jiangsu	1	1				
	Fujian	2	2				
	Guangxi	7 (5)	5 (3)	2 (2)			
Cluster 3		3 (1)	1	1 (1)		1	
	Fujian	1 (1)		1 (1)			
	Guizhou	1	1				
	Guangdong	1				1	
MSM-related clusters							
Cluster 4		15 (1)	1	1 (1)	12	1	
	Beijing	11			11		4a (9), 4b (2)
	Tianjin	1			1		4a (1)
	Jiangsu	2 (1)	1	1 (1)			4b(2)
	Hunan	1				1	4b (1)
Cluster 5		9	3		6		
	Liaoning	3	1		2		
	Beijing	2			2		
	Jilin	4	2		2		
Southern Hetero-clusters							
Cluster 6		5 (3)	5 (3)				
	Fujian	5 (3)	5 (3)				
Cluster 7		3	3				
	Yunnan	3	3				
Ungrouped							
		18 (8)	14 (8)	2		2	
	Fujian	9 (8)	8 (8)			1	
	Yunnan	6	4	2			
	Guizhou	1				1	
	Guangdong	2	2				
	Total	102 (27)	58 (21)	20 (6)	18	6	

[a]The numbers of CRF01_AE NFLG sequences downloaded from the Los Alamos HIV sequence database are shown in the parenthesis.
[b]Provinces are aligned from the north to the south by their geographic locations (See also Fig. 2).
[c]Risk groups: Hetero, heterosexual; n/a, not available.

(Guangxi, $n=7$; Fujian $n=2$) and eastern China ($n=1$) (Fig. 1) was placed within a large phylogenetic group comprised of CRF01_AE strains from Vietnam (designated as VN/CN cluster) (Fig. 1). As previously reported by Liao *et al.* [7], CRF01_AE strains from Vietnam can be divided into three groups that are defined by risk groups and geographic regions (Fig. 1): CRF01_AE variants circulating among heterosexuals and IDUs in southern Vietnam (SV Hetero and SV IDU clusters, respectively), and among IDUs in northern Vietnam and China's Guangxi province (NV/GX IDU cluster). Here, Cluster 2 entirely encompasses the sequences from the NV/GX IDU cluster, and is closely related to the SV IDU cluster, with which it is grouped together with a bootstrap score of 100%.

On the contrary, clusters 4 ($n=15$) and 5 ($n=9$) appeared to be associated with the epidemic among MSM in China. In fact, all CRF01_AE sequences identified from MSM in China in the present study ($n=18$) belonged to either cluster 4 ($n=12$) or cluster 5 ($n=6$) (Table 2 and Fig. 1). Cluster 4 was comprised of two subclusters, which are labeled as 4a ($n=10$) and 4b ($n=5$) (Fig. 1). Subcluster 4a was exclusively from MSM in northern China [Beijing ($n=9$); Tianjin ($n=1$)], whereas subcluster 4b contained CRF01_AE strains from non-MSM populations (or risk factor unknown) from eastern and southern provinces (three of five) (Fig. 1). Cluster 5 strains ($n=9$) were found among MSM ($n=6$) and heterosexuals ($n=3$) in Beijing ($n=2$) and other northeastern provinces [Jilin ($n=4$) and Liaoning ($n=3$)].

Clusters 6 ($n=5$) and 7 ($n=3$) are province-specific clusters detected only among heterosexuals in Fujian and Yunnan provinces, respectively (Table 2 and Fig. 1).

Finally, the remaining 18 HIV-1 strains from China were intermingled with CRF01_AE sequences of Thai origin (designated 'ungrouped'; Table 2) and were all from southern provinces [Fujian ($n=9$); Guangdong ($n=2$); Guizhou ($n=1$); Yunnan ($n=6$)] (Fig. 1 and Table 2). The geographic and risk group distribution of the CRF01_AE clusters identified in this study are illustrated in Fig. 2 (see also Table 2).

Evolutionary characterization of CRF01_AE clusters in China

To explore the timeline of the emergence of CRF01_AE in China, we estimated the time of the most recent common ancestor (tMRCA) of each cluster using a relaxed molecular clock approach [21]. Analyses were performed in BEAST v1.6.1, using three different subgenomic regions [*gag* (HXB2: 790–2292nt); *pol* (HXB2: 2085–5096nt); *env* (HXB2: 6225 –8795nt)], under an uncorrelated lognormal relaxed clock model with a GTR+Γ4 substitution model and a Bayesian skyline demographic model. The estimated evolutionary rates were 5.05 $(4.46–5.65)\times10^{-3}$, 3.25 $(2.94–3.58)\times10^{-3}$, and 6.59 $(5.96–7.21)\times10^{-3}$ substitutions/site per year for the *gag*, *pol*, and *env* regions, respectively (numbers in parenthesis show the 95% highest posterior density for each estimate). The estimates and 95% credible regions of the tMRCAs of each CRF01_AE lineage are depicted in Fig. 3 (see also supplemental Table S2, http://links.lww.com/QAD/A341). The tMRCAs of lineages from central Africa and Thailand were estimated to be in the mid-1970s and mid-1980s, respectively, in accordance with the literature. The tMRCAs of the seven Chinese CRF01_AE lineages fell within in a narrow time range dated from early to late

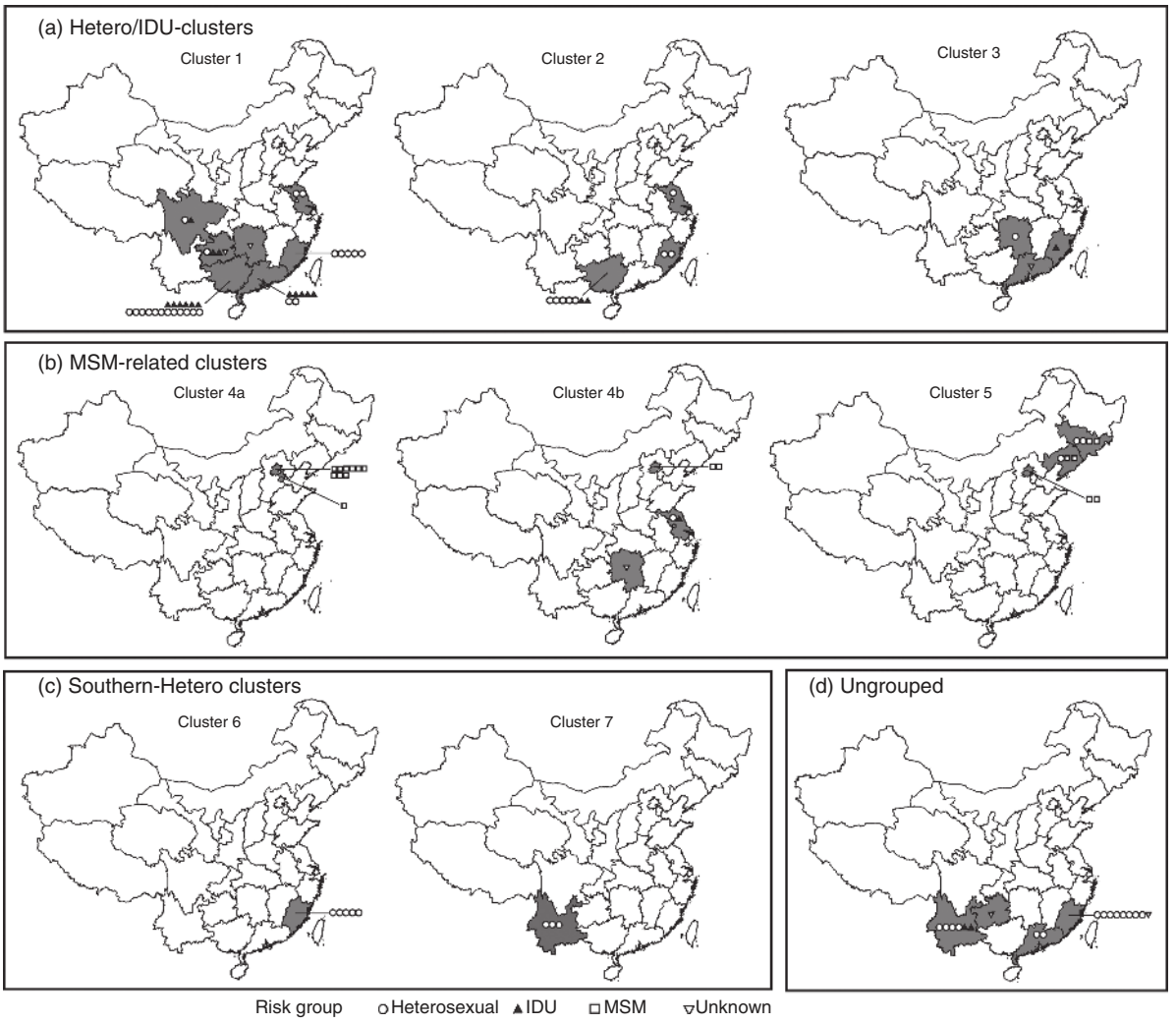

Fig. 2. Geographic and risk group distributions of seven distinct phylogenetic clusters of the CRF01_AE epidemic in China. Clusters 1, 2 and 3 are associated with the epidemics among heterosexuals and IDUs in China (a); Clusters 4 (4a and 4b) and 5 may be associated primarily with MSM epidemics (b); Clusters 6 and 7 are associated with the province-specific heterosexual epidemics in the south (c); Ungrouped (strains in the large radiation of Thai CRF01_AE) (d). Risk categories are shown in different symbols: Heterosexual (○); IDU (▲); MSM (□); and unknown (▽).

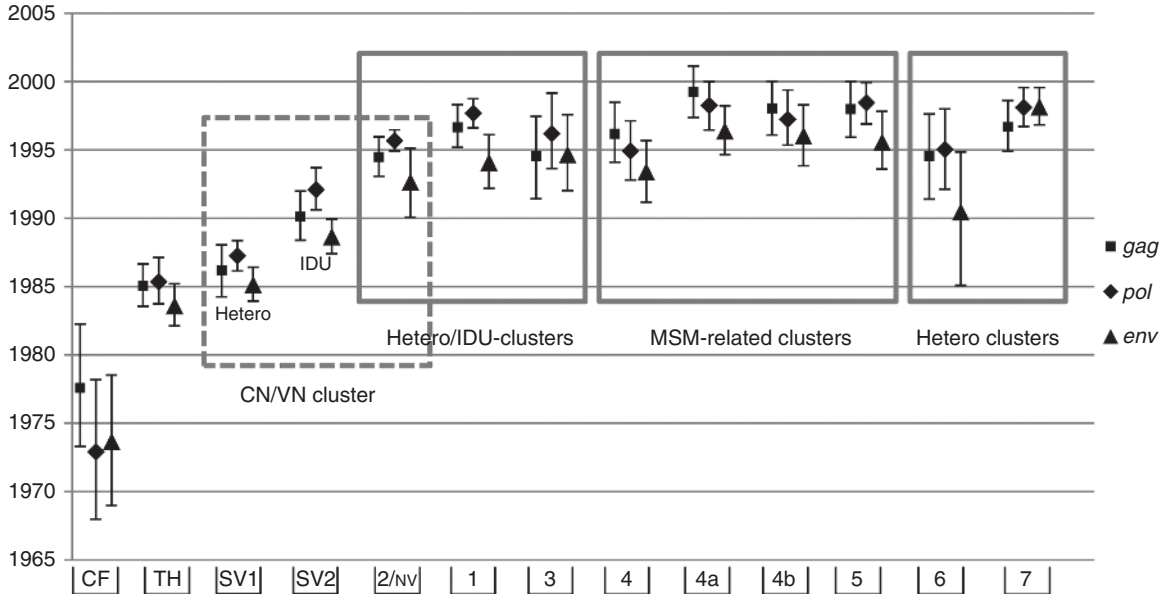

Fig. 3. Estimated tMRCAs of CRF01_AE clusters identified in China. Estimated dates of origin of the seven unique CRF01_AE phylogenetic clusters identified in China (labeled 1 through 7 in square brackets; cluster 4 is subdivided into 4a and 4b), in comparison with CRF01_AE strains from Central African Republic (labeled CF), Thailand (labeled TH) and Vietnam. The clusters from heterosexuals and IDUs in southern Vietnam are labeled with SV1 and SV2, respectively. China's cluster 2 and the northern Vietnam cluster are labeled with 2/NV. Molecular clock analyses were performed using BEAST v1.6.1 (see Materials and Methods). Error bars represent the 95% higher probability density credible regions for each estimate. For each cluster, estimates were obtained using either the *gag* (square), *pol* (diamond), or *env* (triangle) viral sequence (see also supplemental Table S1, http://links.lww.com/QAD/A341).

1990s (Fig. 3; supplemental Table S2, http://links.lww.com/QAD/A341). In general, the substitution and evolutionary models and genomic regions used for the analyses had no significant impact on tMRCA estimates (supplemental Table S2, http://links.lww.com/QAD/A341).

Discussion

In this study, we identified seven independent lineages of HIV-1 CRF01_AE strains circulating in China, the origins of which all date to the mid-to-late 1990s (Fig. 1, Fig. 3). Each cluster appears to have a unique role in China's CRF01_AE HIV-1 epidemic, with distinct associations with particular risk groups and geographic regions (Fig. 2, Table 2). Clusters 1 and 3 were found among heterosexuals and IDUs in mostly southern provinces; the high prevalence of CRF01_AE in these regions is supported by former surveys on southern China [24,25]. Cluster 2, which contained strains circulating among IDUs in Guangxi province, as well as northern Vietnam, was most likely a descendant of strains found among IDUs in southern Vietnam, which clustered closely [7]. Clusters 4 and 5 were highly represented by the more recently emerged transmission among MSM (13 of 15 and 6 of 9, respectively; Table 2) in Beijing and other northern provinces. Clusters 6 and 7 were small, region-specific clusters found only among heterosexuals in Fujian and Yunnan provinces, respectively. Together, these clusters demonstrate the exceptional diversity of the HIV-1 CRF01_AE epidemic in China.

Two observations about the origin and spread of China's CRF01_AE epidemic can be inferred from the phylogenetic structure and transmission patterns. First, although both CRF01_AE and subtype B′ strains entered China following earlier epidemics in nearby Southeast Asian countries [1–3,7–11,26–29], including Thailand, their transmission through distinct risk populations resulted in highly dissimilar geographic and demographic distributions in China. Previous study has shown that following border entry through IDU populations in Yunnan, subtype B′ HIV-1 strains were introduced to plasma donors in central China and consequently a single lineage was amplified throughout the region and the rest of the country [17]. As of 2011, all subtype B′ lineages in circulation in China could be traced to this common ancestor [17]. In contrast, our findings here concerning CRF01_AE showed multiple independent and approximately concurrent introductions into China, followed by population-specific and region-specific transmission through predominantly sexual routes (Fig. 2). These different transmission patterns of subtype B′ and CRF01 AE HIV-1 underscore the complexity of sexual transmission of HIV-1, which involves a more diffuse risk population and is, thus, much more difficult to

control than blood transmission routes. As CRF01_AE becomes the most prevalent strain among heterosexual transmitters and MSM in China [13–16,30], it is likely that we will observe continued evolution and diversification within and across the seven major viral lineages.

Second, all CRF01_AE clusters identified here were estimated to have originated in China during a short window period from mid to late 1990s (Fig. 3). Although previous evidences have implicated the entry of CRF01_AE into China with concurrent epidemics in neighboring countries [7–10], there has been no comprehensive analysis of the subtype's origin in China due to limited samples of those studies. Here, we observe that all major clusters and ungrouped sequences, except for cluster 2, bear direct phylogenetic relationships to sequences from Thailand. Demographic data on travel supports this observation: this was likely a result of vastly increased international travel to and from China, starting in the early 1990s when the Chinese government began relaxing border restrictions to Chinese citizens for unofficial international travel (so called 'Free travel policy') as part of the open door and economic reform policy of the early 1990s [31]. Thailand has been the top destination of outbound tourism for Chinese in the first decade of 'China's free travel policy', as it was the first country that Chinese citizens were allowed to visit (Table 3) [3,32]. This timing coincides with Thailand's explosive HIV-1 epidemic which peaked in mid-1990s and driven by the country's sex industry through heterosexual transmission of CRF01_AE strains. After the Thai government successfully initiated the '100% Condom Program', a rapid increase in condom usage rate was followed by significant decreases in HIV incidence among both direct and indirect CSWs in the country beginning from the late 1990s (Table 3). It is likely that the few years between the opening up of Chinese travel to Thailand and Thailand's effective reduction of HIV incidence served as a short window for the introduction of multiple CRF01_AE strains from Thailand into China.

Additionally, the 'ungrouped' CRF01_AE strains in our phylogenetic analysis (18 of 102, 17.6%) were exclusively from southern provinces in China: Fujian, Yunnan, Guizhou, and Guangdong (Table 2, Fig. 2). Although we cannot formally rule out sampling bias (despite of the relatively large sample size of NFLGs, $n = 102$;), this implicates the earliest entries of CRF01_AE strains into China from Southeast Asia to have taken place in individuals and populations in southern China, especially Fujian, Yunnan and Guangdong provinces, where the earliest epidemics among heterosexuals in China have been reported.

In summary, our results show, for the first time, that the CRF01_AE epidemic in China is remarkably complex,

Table 3. Outbound tourists from China by year, compared with HIV prevalence in Thailand.

	Outbound tourism[a]				HIV infections in Thailand[c]	
Year	Total number	Business travel	Private travel	Rank of Thailand as a travel destination[b]	Prevalence among brothel-based fCSWs	Prevalence among Indirect fCSWs[d]
1990[e]	–	–	–	–	16.83%	4.34%
1991[f]	–	–	–	–	22.72%	6.32%
1992[g]	292.9	180.9	111.9	1	28.17%	7.07%
1993	374.0	227.4	146.6	1	29.52%	9.25%
1994	373.4	209.1	164.2	1	33.54%	10.13%
1995	452.1	246.7	205.4	1	–	18.81%
1996	506.1	264.7	241.4	1	28.34%	11.79%
1997	532.4	288.4	244.0	1	25.70%	9.96%
1998	842.6	523.5	319.0	1	22.10%	7.54%
1999	923.2	496.6	426.6	1	17.84%	7.45%
2000	1047.3	484.2	563.1	1	17.61%	6.59%
2001	1213.3	518.8	694.5	1	16.20%	5.43%
2002	1660.2	654.1	1006.1	3	12.89%	4.56%
2003	2022.2	541.1	1481.1	5	10.67%	4.67%
2004	2885.3	587.4	2297.9	6	9.97%	4.52%
2005	3102.6	588.6	2514.0	5	8.19%	3.78%
2006	3452.4	572.4	2879.9	3	6.06%	2.93%
2007	4095.4	603.0	3492.4	5	6.77%	3.27%
2008	4584.4	571.3	4013.1	7	5.57%	2.88%
2009	4763.6	587.9	4221.0	8	3.53%	2.20%
2010	5738.7	544.7	5150.9	7	–	–

[a]10 000 person-times per year. Data from 'The Yearbook of China Tourism Statistics' (1990–2011).
[b]Thailand was the leading travel destination for Chinese citizen in 1992–2001, except Hong Kong, Macau and Taiwan.
[c]Data from Thailand Bureau of Epidemiology (http://www.boe.moph.go.th/).
[d]Indirect fCSWs refers to a female sex worker whose service comes from karaoke bars and internet.
[e]Chinese government began allowing unsponsored international travel ('Free travel') for Chinese citizens to Thailand in 1990.
[f]'Free travel policy' extended to Singapore and Malaysia in addition to Thailand,.
[g]Since then 'Free travel' policy were extended to more and more countries.

comprising of multiple genetically distinct lineages that were independently introduced into China during the mid-to-late 1990s and subsequently spread into different risk groups and geographic regions. Further study and surveillance of China's CRF01_AE lineages will provide knowledge on their increasing transmission and continued evolution, which will help to inform effective intervention programs.

Acknowledgements

The authors would like to thank Yao Yang and Jing Wei for technical support, Dr Punnee Pitisuttithum of Mahidol University for assistance in acquiring official Bureau of Epidemiology data records on HIV prevalence among Thailand's sex workers.

The authors would like to thank the Group for HIV Molecular Epidemiologic Survey. Contributing members of the Group [name of contributor (facility of the contributor)]: Hongyan Lu (Beijing CDC); Pingping Yan (Fujian CDC); Peng Lin (Guangdong CDC); Shen Zhiyong, Shujia Liang (Guangxi CDC); Xianguang Sun (Guizhou CDC); Xi Chen, Jianmei He (Hunan CDC); Haitao Yang, Xiaoqin Xu (Jiangsu CDC); Xiangdong Meng, Xihui Zang (Jilin CDC); Fengxia Jiang (Liaoning CDC); Guangming Qin, Shu Liang (Sichuan CDC); Xiaoke Zhu, Minna Zheng (Tianjin CDC); Yanling Ma, Manhong Jia (Yunnan CDC).

Y.S., X.H. and Y.F. conceived and designed the study. Y.F., F.L., X.L., Q.W., T.L., and X.H. performed the experiments and analyzed the data. Y.F., J.H.H., Y.T. and Y.S. drafted the manuscript. Y.S., X.H., Y.R., H.X., and O.P. interpreted data and provided critical review. All authors reviewed and approved the final manuscript.

This study was supported by the National Science and Technology Major Project for Infectious Diseases Control and Prevention (2008ZX10001–004, 2012ZX10001–002), and 2012ZX10001 - 008. National Natural Science Foundation of China (81261120379), NIH Foundation (1R01AI094562-01). International Cooperative Grant (2009DFB30420) and SKLID Development Grant (2012SKLID103).

Conflicts of interest

All authors declare that they have no conflicts of interest.

References

1. McCutchan FE, Hegerich PA, Brennan TP, Phanuphak P, Singharaj P, Jugsudee A, *et al*. **Genetic variants of HIV-1 in Thailand.** *AIDS Res Hum Retroviruses* 1992; **8**:1887–1895.
2. Carr JK, Salminen MO, Koch C, Gotte D, Artenstein AW, Hegerich PA, *et al*. **Full-length sequence and mosaic structure of a human immunodeficiency virus type 1 isolate from Thailand.** *J Virol* 1996; **70**:5935–5943.
3. Nelson KE, Celentano DD, Suprasert S, Wright N, Eiumtrakul S, Tulvatana S, *et al*. **Risk factors for HIV infection among young adult men in northern Thailand.** *JAMA* 1993; **270**:955–960.
4. Gao F, Robertson DL, Morrison SG, Hui H, Craig S, Decker J, *et al*. **The heterosexual human immunodeficiency virus type 1 epidemic in Thailand is caused by an intersubtype (A/E) recombinant of African origin.** *J Virol* 1996; **70**:7013–7029.
5. Robertson DL, Sharp PM, McCutchan FE, Hahn BH. **Recombination in HIV-1.** *Nature* 1995; **374**:124–126.
6. McCutchan FE, Artenstein AW, Sanders-Buell E, Salminen MO, Carr JK, Mascola JR, *et al*. **Diversity of the envelope glycoprotein among human immunodeficiency virus type 1 isolates of clade E from Asia and Africa.** *J Virol* 1996; **70**:3331–3338.
7. Liao H, Tee KK, Hase S, Uenishi R, Li XJ, Kusagawa S, *et al*. **Phylodynamic analysis of the dissemination of HIV-1 CRF01_AE in Vietnam.** *Virology* 2009; **391**:51–56.
8. Chen J, Young NL, Subbarao S, Warachit P, Saguanwongse S, Wongsheree S, *et al*. **HIV type 1 subtypes in Guangxi Province, China, 1996.** *AIDS Res Hum Retroviruses* 1999; **15**:81–84.
9. Piyasirisilp S, McCutchan FE, Carr JK, Sanders-Buell E, Liu W, Chen J, *et al*. **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* 2000; **74**:11286–11295.
10. Yu XF, Chen J, Shao Y, Beyrer C, Lai S. **Two subtypes of HIV-1 among injection-drug users in southern China.** *Lancet* 1998; **351**:1250.
11. Cheng H, Zhang J, Capizzi J, Young NL, Mastro TD. **HIV-1 subtype E in Yunnan, China.** *Lancet* 1994; **344**:953–954.
12. Xing H, Pan PL, Su L, Fan XJ, Feng Y, Qiang LY, *et al*. **[Molecular Epidemioiogical Study of a HIV-1 Strain of subtype E in China between 1996 and 1998].** *Zhongguo Xing Bing Ai Zi Bing Fang Zhi* 2002; **8**:200–203.
13. Xing H, Liang H, Wan ZY, Chen X, Wei M, Ma PF, *et al*. **[Distribution of recombinant human immunodeficiency virus type-1 CRF01_AE strains in China and its sequence variations in the env V3-C3 region].** *Zhonghua Yu Fang Yi Xue Za Zhi* 2004; **38**:300–304.
14. He X, Xing H, Ruan Y, Hong K, Cheng C, Hu Y, *et al*. **A comprehensive mapping of HIV-1 genotypes in various risk groups and regions across China based on a nationwide molecular epidemiologic survey.** *PLoS One* 2012; **7**:e47289.
15. Zhao B, Han X, Dai D, Liu J, Ding H, Xu J, *et al*. **New trends of primary drug resistance among HIV type 1-infected men who have sex with men in Liaoning Province, China.** *AIDS Res Hum Retroviruses* 2011; **27**:1047–1053.
16. Zhang X, Li S, Li X, Xu J, Li D, Ruan Y, *et al*. **Characterization of HIV-1 subtypes and viral antiretroviral drug resistance in men who have sex with men in Beijing, China.** *AIDS* 2007; **21 (Suppl 8)**:S59–S65.
17. Li Z, He X, Wang Z, Xing H, Li F, Yang Y, *et al*. **Tracing the origin and history of HIV-1 subtype B′ epidemic by near full-length genome analyses.** *AIDS* 2012; **26**:877–884.
18. Hall TA. **BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT.** *Nucleic Acids Symp Ser* 1999; **41**:95–98.
19. Posada D, Crandall KA. **MODELTEST: testing the model of DNA substitution.** *Bioinformatics* 1998; **14**:817–818.
20. Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O. **New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0.** *Syst Biol* 2010; **59**:307–321.
21. Drummond AJ, Ho SY, Phillips MJ, Rambaut A. **Relaxed phylogenetics and dating with confidence.** *PLoS Biol* 2006; **4**:e88.
22. Drummond AJ, Nicholls GK, Rodrigo AG, Solomon W. **Estimating mutation parameters, population history and genealogy simultaneously from temporally spaced sequence data.** *Genetics* 2002; **161**:1307–1320.
23. Drummond AJ, Rambaut A. **BEAST: Bayesian evolutionary analysis by sampling trees.** *BMC Evol Biol* 2007; **7**:214.

24. Cheng CL, Feng Y, He X, Lin P, Liang SJ, Yi ZQ, *et al*. **[Genetic characteristics of HIV-1 CRF01_AE strains in four provinces, southern China].** *Zhonghua Liu Xing Bing Xue Za Zhi* 2009; **30**:720–725.
25. Zeng H, Sun Z, Liang S, Li L, Jiang Y, Liu W, *et al*. **Emergence of a New HIV Type 1 CRF01_AE Variant in Guangxi, Southern China.** *AIDS Res Hum Retroviruses* 2012; **28**:1352–1356.
26. Ma Y, Li Z, Zhao SD. **HIV infected people were first identified in intravenous drug users in China.** *Chin J Epidemiol* 1990; **11**:184–185.
27. Ou CY, Takebe Y, Weniger BG, Luo C, Kalish ML, Auwanit W, *et al*. **Independent introduction of two major HIV-1 genotypes into distinct high-risk populations in Thailand.** *Lancet* 1993; **341**:1171–1174.
28. Shao Y, Chen Z, Wang B, Zeng Y, Zhao SD, Zhang ZR. **Isolation of viruses from HIV infected individuals in Yunnan.** *Chin J Epidemiol* 1991; **12**:129.
29. Shao Y, Zhao QB, Wang B, Chen Z, Su L, Zeng Y, *et al*. **Sequence analysis of HIV env genes among HIV infected drug injecting users in Dehong epidemic area of Yunnan province, China.** *Chin J Virol* 1994; **10**:291–299.
30. Li L, Lu X, Li H, Chen L, Wang Z, Liu Y, *et al*. **High genetic diversity of HIV-1 was found in men who have sex with men in Shijiazhuang, China.** *Infect Genet Evol* 2011; **11**:1487–1492.
31. *The yearbook of China tourism statistics*. Beijing: National Tourism Administration of the P.R.C. 1978–2011.
32. Rojanapithayakorn W, Hanenberg R. **The 100% condom program in Thailand.** *AIDS* 1996; **10**:1–7.