To explore the temporal genetic variation of human immunodeficiency virus type 1 CRF07_BC and reconstruct its epidemic in Xinjiang, China, we studied 216 C2–V4 fragments of env genes sampled from 1996 to 2008. Phylogenetic analysis indicates that the viruses prevailing in Xinjiang form a large monophyletic cluster and may have originated from a common ancestor. The epidemic in Xinjiang was probably established around 1995 (95% confidence interval, 1994–1996). We noted an increased diversity of CRF07_BC over time, with a rapid evolutionary rate we estimated to be $8.3 \times 10^{-3}$ substitutions per site per year in the env gene. After 5–6 years of the epidemic (1997–2002), the transmission rate of CRF07_BC in Xinjiang slowed down, although CRF07_BC infection remained at a high prevalence.

In China, the local epidemic of HIV-1 subtype B’ (Thai B) in 1989 and C in 1992 among intravenous drug users (IDUs) in Yunnan triggered the severe HIV-1 epidemic in China (Luo et al., 1995; Ma et al., 1990; Shao et al., 1994; Wu et al., 2007; Zhang et al., 2006; Zheng et al., 1994). CRF07_BC, a recombinant derived from subtypes B’ and C, was first identified among IDUs in Sichuan and Xinjiang in 1996 (Fig. 1) (Shao et al., 1999; Su et al., 2000). HIV-1 CRF07_BC has predominated in Xinjiang for over twelve years (Shao et al., 1999; Su et al., 2000), mainly transmitted in IDUs (over 80%). Results from continuous HIV subtype surveillance indicate that CRF07_BC accounted for over 90% of HIV-1 infections in Xinjiang in the past 12 years (Liu et al., 2008; Meng et al., 2008) and almost one quarter of infections in mainland China (unpublished data). In 2003, CRF07_BC was identified among IDUs in Taiwan (Chen et al., 2006), with a prevalence of 15–20% amongst over 60 000 IDUs (Chen & Kuo, 2007). Although some investigations of the genetic characteristics of CRF07_BC have been reported (Liu et al., 2008; Meng et al., 2007; Song et al., 2007; Su et al., 2000), little is known about the transmission network, the genetic dynamics of CRF07_BC, and even its origin in China.

We therefore analysed the genetic dynamics and phylogenetic relationships of longitudinally sampled env genes of CRF07_BC. Blood samples were collected from HIV-1-infected individuals in Xinjiang, spanning 1996 to 2008. The C2–V4 fragments of the env gene (nucleotide positions 7080–7395 of HXB2) were amplified from DNA extracted from whole blood samples collected before 2006 or RNA extracted from plasma samples from 2006 to 2008. The C2–V4 sequences of CRF07_BC without discernible recombination, indels ending in frameshifts, or distorted evolutionary process (e.g. hypermutation) were included in the phylogenetic analysis. In total, 216 env gene sequences were obtained (GenBank accession nos FJ875724–FJ875939), including 74 from Urumqi (central Xinjiang), 116 from Yili (western Xinjiang), 3 from Karamay (north Xinjiang), and 17, 5 and 1 from Aksu, Kashkar and Hetian, respectively, in south Xinjiang (Fig. 1). Epidemiologically, 176 out of 216 cases were infected by intravenous drug use (IDU), 11 by heterosexual contact, 17 by IDU or heterosexual route, and the transmission routes of 12 infected individuals were not available. The demographic information and sampling year are summarized in Table 1.

1Present address: Medical Laboratory of Shenzhen International Travel Healthcare Center, Shenzhen 518033, PR China.

The GenBank/EMBL/DDBJ accession numbers of the sequences reported in this study are FJ875724–FJ875939.

Supplementary material is available with the online version of this paper.
All the sequences were cleaned and codon-aligned using Sequencher and BioEdit software, respectively. Neighbour-joining (NJ) phylogeny constructed by MEGA 4.0 software was used to determine subtypes. The maximum-likelihood (ML) phylogeny of 216 env sequences in Xinjiang and 19 subtyping reference sequences selected from Los Alamos HIV sequence database was constructed by PAUP 4.0beta and a maximum-likelihood code with a HKY (Hasegawa et al., 1985) plus gamma model with variable rates among sites (PAUP command line kindly offered by Dr Oliver G. Pybus). The 216 env sequences were randomly distributed and formed a unique cluster surrounding 97CN001 (GenBank accession no. AF286226), a reference sequence derived from the isolate in Xinjiang in 1997, either in the NJ or ML phylogenetic trees (Supplementary Figs S1 and S2, available in JGV Online). The sequences were unanimously subtyped as CRF07_BC, with the assistance of corresponding gag or tat genes. The sequences from any group (sampling year) were intermingled with high genetic homology, regardless of geographical region or sampling year biases. Therefore, CRF07_BC strains may spread through ongoing local transmissions and account for the HIV-1 epidemic in Xinjiang during 1996–2008, probably as a consequence of a founder effect. The CRF07_BC strains in Xinjiang were postulated to originate from a common lineage, forming the large monophyletic cluster (Fig. 2a).

Bayesian methods were utilized for the inference, which allowed for the estimation of phylogeny and divergence times of CRF07_BC in Xinjiang under an uncorrelated exponential relaxed molecular clock model, using BEAST v1.4.8 (Drummond & Rambaut, 2007). All analyses were

---

**Table 1.** The demographic information and sampling year of the cases infected with CRF07_BC in Xinjiang

<table>
<thead>
<tr>
<th>Year</th>
<th>Urumqi</th>
<th>Yili</th>
<th>Karamay</th>
<th>Aksu</th>
<th>Kashkar</th>
<th>Hetian</th>
<th>Number of sequences</th>
</tr>
</thead>
<tbody>
<tr>
<td>1996</td>
<td>2</td>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>6</td>
</tr>
<tr>
<td>1997</td>
<td>6</td>
<td>12</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td></td>
<td>24</td>
</tr>
<tr>
<td>1998</td>
<td>7</td>
<td>4</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td>12</td>
</tr>
<tr>
<td>1999</td>
<td>8</td>
<td>11</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td>20</td>
</tr>
<tr>
<td>2000</td>
<td>4</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td></td>
<td></td>
<td>11</td>
</tr>
<tr>
<td>2001</td>
<td>4</td>
<td>3</td>
<td></td>
<td>7</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2002</td>
<td>7</td>
<td>17</td>
<td>6</td>
<td>3</td>
<td></td>
<td></td>
<td>33</td>
</tr>
<tr>
<td>2003</td>
<td>6</td>
<td></td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td>7</td>
</tr>
<tr>
<td>2004</td>
<td>8</td>
<td>30</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>38</td>
</tr>
<tr>
<td>2005</td>
<td>3</td>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>9</td>
</tr>
<tr>
<td>2007</td>
<td>23</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>23</td>
</tr>
<tr>
<td>2008</td>
<td>26</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>26</td>
</tr>
</tbody>
</table>
performed under an HKY nucleotide substitution model with heterogeneity among sites modelled with a *gamma* distribution and invariant sites. Several demographic models (constant population, exponential growth, logistic growth) were used for the inference, including a Bayesian skyline plot coalescent tree prior (Drummond *et al.*, 2005). Bayes factors were used to evaluate the marginal likelihoods for each model (Salemi *et al.*, 2001). The best-fit model incorporated an exponential population size demographic model to estimate the time to the most recent common ancestor (*t*~MRCA~) as 14.2 years (95% highest-posterior-density [HPD] 13.0–15.4). Another best-fit model with a general, non-parametric prior (the Bayesian skyline plot tree prior) indicated that the CRF07_BC epidemic in Xinjiang was established approximately 13.3 years ago (95% HPD, 12.4–14.5). Therefore, CRF07_BC may have been introduced into Xinjiang around 1995 (95% confidence interval, 1994–1996) or before (Fig. 2a, b), heralding the epidemic of HIV-1 infections through ongoing local transmissions. Epidemiologically, according to the national HIV sentinel surveillance, the first HIV-1 seropositive case in Xinjiang was detected amongst 404 IDUs in 1995 (Department of Disease Control, Ministry of Health, PR China, 1996). Given that more than 90% HIV-1/AIDS cases in Xinjiang were caused by CRF07_BC (Liu *et al.*, 2008; Meng *et al.*, 2008; unpublished data), especially amongst IDUs, we believe that CRF07_BC was the pathogen infecting this first HIV-1 case. Furthermore, the first CRF07_BC epidemic in Xinjiang was reported in early 1996 (Bai *et al.*, 1997; Shao *et al.*, 1999). All aforementioned evidence supports our results about the origin of the CRF07_BC epidemic in Xinjiang.

Tee *et al.* (2008) used Bayesian inference, based on 22 gag gene fragments of CRF07_BC obtained from Yunnan, Xinjiang and Liaoning, to deduce that the origin of CRF07_BC was around 1993 in Yunnan, China. Two years later, the virus was transmitted into Xinjiang through heroin trafficking, and first identified there 1 year later. Hence, the present-day data support the hypothesis of the CRF07_BC origin in China and its transmission network along the heroin trafficking route from Yunnan (around 1993) or Sichuan via Ningxia to Xinjiang (around 1995) by Shao *et al.* (1999) (Fig. 1).

A Bayesian skyline plot model deciphered the past population dynamics of the CRF07_BC epidemic in Xinjiang through time. The onset of the CRF07_BC epidemic may have begun approximately 13 years ago (around 1995) in Xinjiang, at a smaller effective population
size (EPS) and a lower transmission rate (Fig. 2b). A fast transmission rate of CRF07_BC was observed between 1997 and 2002, with CRF07_BC-infected cases (EPS) mounting sharply. Epidemiological data indicated that there was an increasing trend of HIV-1-infected cases across Xinjiang from 1996. The HIV-1-positive rate in spouses or sexual partners of IDUs was higher (13.5 %), which suggested that CRF07_BC was spreading to the general population through heterosexual contact (Ni et al., 2004). From 2003, the transmission rate of CRF07_BC amongst the high risk population slowed down, which indicated the infected population got closer to saturation. However, a high prevalence of CRF07_BC infections was maintained in Xinjiang (Fig. 2b).

Previous studies indicated that, after transmission, most of the diversity in strains from the donor was lost in the recipient, especially in mother-to-child (Wolinsky et al., 1992) or homosexual transmissions (Edwards et al., 2006), which caused the greater distances and variation among the spreading strains. This is not the case during the transmission via IDU, as close genetic relatives of HIV-1 strains were reported amongst IDU networks (Saad et al., 2006; Shao et al., 1999). In this study, much lower distances (below 0.03) were observed during the initial period of epidemic (from 1996 to 1999) (Fig. 2c), which is consistent with the results obtained previously (Piyasirisilp et al., 2000; Shao et al., 1999). These results indicated that the homogeneous CRF07_BC strains were spreading out across Xinjiang.

During the fast epidemic of HIV-1 amongst IDUs, especially in a smaller IDU network, HIV-1 transmission began shortly after the initial infection. The strains were prone to be transmitted at the first stage of intrahost infection, before the host imposed immune pressure on them (Berry et al., 2007). Therefore, the strains with high genetic homogeneity were transmitted. The Bayesian skyline plot illustrated that fast epidemic of CRF07_BC was observed during this early period in Xinjiang (Fig. 2b), when the viruses were transmitted in a comparatively smaller IDU population. The fast transmission rate pertaining to the particular transmission mode (through IDU) may explain the lower diversity and divergence of the circulating strains. Comparatively, the evolutionary rate of CRF07_BC at this fast epidemic stage was assumed to be very low. As the epidemic went on, CRF07_BC evolved more quickly, generating higher genetic diversity. The genetic distances increased dramatically from year 2000. At present, the distances within a group (about 0.09) are about four- to fivefold higher than those in the initial stage (Fig. 2c), which suggests genetically heterogeneous CRF07_BC strains were transmitted amongst IDUs after the year 2000. Likewise, env genes of CRF07_BC evolved slowly during the initial period, and faster thereafter (Fig. 2c).

The increasing genetic diversity at a population level suggested a high evolutionary rate during the CRF07_BC epidemic. We used a best-fit model of the Bayesian skyline plot to deduce the evolutionary rate (μ) for the env gene C2–V4 of CRF07_BC in Xinjiang covering 1996–2008 as 8.36 × 10⁻² substitutions per site per year (95 % HPD, 6.91–9.87 × 10⁻²) (Fig. 2a, Supplementary Table S1). The rate was about three- to fivefold higher than those of the(env V3 fragment (2.3–6.7 × 10⁻³) or the whole env gene (1.0–1.7 × 10⁻⁵) (Korber et al., 1998, 2000; Salemi et al., 2001). As the CRF07_BC epidemic developed, mainly in the IDU population (over 80 %), the CRF07_BC-infected population approached saturation, as indicated by the Bayesian coalescent skyline plot (Fig. 2b). Correspondingly, the virus expressed high diversification, particularly in the late epidemic. Our observations are well explained by the argument of an inverse relationship between the HIV-1 evolutionary rate and HIV-1 transmission rate (Berry et al., 2007). However, it should be noted that CRF07_BC infections remained highly prevalent across Xinjiang in recent years (Fig. 2b), and there was a trend of transmitting CRF07_BC from the high-risk to the general population via heterosexual contact. More efforts and actions should be taken to halt CRF07_BC transmission in Xinjiang.

In conclusion, CRF07_BC was transmitted from Yunnan (Shao et al., 1999; Tee et al., 2008) into Xinjiang around 1995, and accounted for most of the HIV infections there through ongoing local transmissions. We highlighted that after 5–6 years of this fast epidemic, the transmission rate of CRF07_BC across Xinjiang slowed down, and the circulating strains exhibited higher genetic diversity and divergence. The CRF07_BC epidemic was saturated and simultaneously the virus evolved at a higher rate. However, CRF07_BC infections remain highly prevalent in Xinjiang, and intervention measures should be reinforced to control its spread.

Acknowledgements

We thank the staff at the Center for Disease Control and Prevention of Xinjiang Uygur Autonomous Region and the Center for Disease Control and Prevention of Yili for their contributions in sample collection. We also thank Dr Oliver G. Pybus from the Department of Zoology, University of Oxford, UK, and Dr James I. Mullins from the University of Washington, WA, USA, for their kind instructions and suggestions in the evolutionary analyses and inference. This work was supported by 863 Program (2006 AA02Z418) from the Ministry of Science and Technology of China. It was also supported by the National Natural Science Foundation of China (30671847) and National Major Projects for Infectious Diseases Control and Prevention (2008ZX1001-004 and 2008ZX1001-010) from the Ministry of Health of China.

References


human immunodeficiency virus type 1 (HIV-1) pandemic: the evolutionary rate of HIV-1 slows down when the epidemic rate increases. *J Virol* 81, 10625–10635.


**Piyasirisilp, S., McCutchan, F. E., Carr, J. K., Sanders-Buell, E., Liu, W., Chen, J., Wagner, R., Wolf, H., Shao, Y. M. & other authors (2000).** A recent outbreak of human immunodeficiency virus type 1 infection in southern China was initiated by two highly homogen- eous, geographically separated strains, circulating recombinant form AE and a novel BC recombinant. *J Virol* 74, 11286–11295.


