VIRUS EVOLUTION

Virus Evolution, 2021, **7(2)**, 1–8

DOI: https://doi.org/10.1093/ve/veab094

Advance access publication date: 9 November 2021

Research Article

Independent epidemic patterns of HIV-1 CRF01_AE lineages driven by mobile population in Shenzhen, an immigrant city of China

Minghui An,[1,2] Chenli Zheng,[3] Hao Li,[3] Lin Chen,[3] Zhengrong Yang,[3] Yongxia Gan,[3] Xiaoxu Han,[1,2,*] Jin Zhao,[3,*] and Hong Shang[1,2,†,*]

[1]NHC Key Laboratory of AIDS Immunology (China Medical University), National Clinical Research Center for Laboratory Medicine, The First Affiliated Hospital of China Medical University, No 155, Nanjing North Street, Shenyang, Liaoning 110001, China, [2]Key Laboratory of AIDS Immunology, Chinese Academy of Medical Sciences, No 155, Nanjing North Street, Shenyang, Liaoning 110001, China and [3]Shenzhen Center for Disease Control and Prevention, No 8, Longyuan Road, Shenzhen, Guangdong 518055, China

[†]https://orcid.org/0000-0001-5333-8943

*Corresponding authors: E-mail: hongshang100@hotmail.com; szhaojin@gmail.com; xxhan@yeah.net

Abstract

Shenzhen, a city with >12 million migrant population, may play a key role in the spread of human immunodeficiency virus (HIV)-1 in China. The transmission dynamics of CRF01_AE, a predominant subtype in Shenzhen, is a good model to characterize the impact of human mobility on HIV-1 epidemic locally and nationally. We used phylodynamic and phylogeographic methods to estimate the viral transmission dynamics and migration trajectory of variable lineages based on 1,423 CRF01_AE sequences in Shenzhen sampled between 2006 and 2015. Eleven lineages of CRF01_AE were detected in Shenzhen. Of those, four main lineages originated during the 1990s. Their basic viral reproduction number (R_0) ranged 1.96–3.92. The effective viral reproduction number (R_e) of two lineages prevalent among heterosexuals/people who inject drugs had reduced <1 at the end of sampling, and the main sources were the intra-provincial immigrants (72 per cent) for one and local residents of Shenzhen (91 per cent) for another. Within two lineages among men who have sex with men (MSM), R_e had been above or close to 1 at the end of sampling, and the immigrants from Jiangxi/Shaanxi and Hubei as sources accounted for 93 per cent and 68 per cent of all viral migration events, respectively. Moreover, no obvious recipients were found throughout the viral migration history for any lineage. Our findings demonstrate that HIV epidemic is declining in Shenzhen, which coincided with the initiation of the interventions during the 2000s. However, the obvious differences of the epidemic patterns between lineages emphasize the importance of further targeting interventions and continued molecular tracing, focusing on high-risk transmission sources among MSM.

Key words: HIV-1 CRF01_AE; phylodynamic; phylogeographic; migration; men who have sex with men.

1. Introduction

By the end of 2019, it was estimated that 38 million people were infected with human immunodeficiency virus (HIV) (UNAIDS 2020). The HIV-1 group M resulted in a pandemic with nine subtypes, 103 circulating recombinant forms (CRFs) and many unique recombinant forms (URFs). Of those, CRF01_AE was one of most prevalent CRFs. It originated from Central Africa in the 1970s and established a broad-scale epidemic in Southeast and East Asia, accounting for 90 per cent of cases in some Asian countries (Angelis et al. 2015; LANL 2019).

In China, CRF01_AE was first found among individuals at risk due to sexual contact and drug injection in the southwest provinces during the early 1990s (Cheng et al. 1994). Since then, it has consistently been one of the most predominant subtypes nationwide (Wang and Zhong 2015). Multiple distinct CRF01_AE lineages were introduced to China and subsequently shaped the epidemic patterns of HIV-1 (An et al. 2012; Abubakar et al. 2013; Feng et al. 2013; Peng et al. 2015; Li et al. 2017; Wang et al. 2017). Currently, at least eight CRF01_AE lineages, originated between 1990 and 2000, circulate in China. Of those, six circulate among heterosexuals and people who inject drugs (PWID), while two among men who have sex with men (MSM) (Li et al. 2017).

Shenzhen, as China's first special economic zone, attracts a large number of immigrants from other regions across China and even overseas. At the end of 2020, Shenzhen reported a population of 17.56 million, and >70 per cent of those do not have local permanent registration (hukou) (Shenzhen-Statistics 2020). It complicates the prevention and control of HIV epidemic via routine interventions in the context of the cross-region viral exchanges. The first patient with acquired immunodeficiency syndrome (AIDS) in Shenzhen was reported in 1992 (Peng 2004). Following a long stagnant status from 1992 to 2003, the number

of HIV diagnoses in Shenzhen began to increase rapidly and the number of newly diagnosed HIV infections were 2,555 in 2015 (Tan et al. 2016; Yang et al. 2018). The active sexual populations with rapid growing rates of HIV infection accounted for >95 per cent of newly diagnosed cases in 2015 (Yang et al. 2018) and since 2014, the number of homosexual HIV infections has surpassed that of heterosexual infections (Tan et al. 2016).

In recent years, CRF07_BC and CRF55_01B exhibited a sharp increase in infections among MSM in Shenzhen; however, CRF01_AE has consistently been one of the most dominant subtypes (Zhao et al. 2012, 2016; Jia et al. 2019). The epidemic of CRF01_AE in Shenzhen is a suitable model that can reflect the impact of human mobility on HIV transmission dynamics and the effectiveness of interventions (Cassels 2020; Ratmann et al. 2020). In this study, we focused on two key points: (1) the impact of population mobility on the epidemic of HIV-1 and (2) the effectiveness of current interventions under the high frequent mobility of population.

2. Materials and methods

2.1 Study dataset and ethical approval

A total of 1,432 HIV-1 CRF01_AE *pol* sequences (HXB2: 2,253–3,311 nt) with the information of the sampling date, household register, sex, age, and risk group were obtained from a local database at the Shenzhen CDC during 2006–2015. Most samples were obtained from research cohorts among MSM, containing two depth-sampling cohorts established in 2009 and 2013. In order to investigate the whole situation of HIV subtypes circulating in Shenzhen, all of the HIV-1 newly diagnosed cases of Shenzhen in 2009 and 2013 were sampled, and the acquired rate of sequences in 2009 was 49 per cent, that of which in 2013 was high to 92 per cent along the development of technique (Supplementary Figure S1). All samples were collected at the time of diagnosis or first follow-up. The usage of viral sequences was approved and governed by the Medical Ethics Committee of the Shenzhen Center for Disease Control and Prevention (CDC).

2.2 Preliminary phylogenetic inference to define CRF01_AE lineages

Reference sequences defining Chinese HIV-1 CRF01_AE lineages and ensuring the topology of phylogenetic tree were downloaded from the HIV database at Los Alamos National Laboratory (http://www.hiv.lanl.gov) (Li et al. 2017). Shenzhen CRF01_AE sequences were aligned with reference sequences in AliView v1.25 and the final alignment was manually edited according to coding positions after the removal of major HIV drug resistance mutation sites described by the International AIDS Society–USA. ModelFinder package in IQ-TREE v1.6.12 was used to determine the most appropriate nucleotide substitution model. The maximum likelihood tree was reconstructed using GTR + I + G4 nucleotide substitution model with 1,000 replicates in IQ-TREE. The branch support was estimated with the approximate likelihood-ratio test and the ultrafast bootstrap method (Nguyen et al. 2015). The lineages were identified using the following standards: (1) approximate likelihood-ratio test value >80 per cent and ultrafast bootstrap value >95 per cent and (2) more than two Shenzhen sequences in one lineage (Bui et al. posting date; Feng et al. 2013; Li et al. 2017; Kostaki et al. 2019; Wilkinson et al. 2019).

For the lineages with >20 sequences, we evaluated the temporal signal between genetic diversity and sampling date in TempEst (http://beast.community/tempest) to identify suitable lineages for the spatiotemporal reconstructions (Rambaut et al. 2016).

2.3 Phylodynamic estimation of HIV-1 CRF01_AE lineages

Due to the depth-sampling, there were much more sequences in 2009 and 2013, which were un-balanced to other sampling years. To avoid the impact of sampling depth on phylodynamic reconstruction, we randomly down-sampled the sequences of these 2 years according to average sequences' acquired rate of 2 years before and after (Supplementary Tables S2 and S3 and Supplementary Figure S2).

The birth-death susceptible-infected-recovered (BDSIR) model in BEAST v 2.6.2 (Kuhnert et al. 2014) was used to estimate epidemiological parameters, with GTR + I + G4 nucleotide substitution model and the relaxed log-normal molecular clock model. In the BDSIR model, the 'S' represents susceptible individuals, the 'I' represents infected individuals, and the 'R' represents the recovered individuals (the cure or death). Although HIV infection cannot be cured, it is thought that patients who achieve virological control under the combination antiretroviral therapy (cART) nearly do not transmit HIV-1. Shenzhen is one of the cities that carried out the prevention programs of HIV/AIDS at the earliest in China. In 1997, Shenzhen started the health education and behavioral intervention among HIV high-risk and key population. In 1998, Shenzhen initiated the antiretroviral therapy for AIDS, much earlier than national policy. By the end 2009, the national average treatment coverage for actively sexual population with HIV in need was 61.7 per cent (Zhang et al. 2011), so we believe that the treatment coverage in Shenzhen would be much higher than national level. Moreover, a conference report indicated that during 2010–14, the percentage of patients achieving virological suppression under cART in Shenzhen was 98.2 per cent (He et al. 2015). Hence, those achieving virological suppression can be removed from the infectious pool and serve as the recovered cases. Therefore, SIR model would also be suitable for this study of HIV in Shenzhen and the 'R' means the individuals who have acquired virological suppression (Kuhnert et al. 2014) .The priors were set as followed: a log-normal (8, 2) for the population size of susceptible individuals, with a 95 per cent confidence interval (CI) of the initial susceptible individuals between 59.1 and 150,000; a log-normal (1.3, 1) for the becoming-non-infectious rate (Dennis et al. 2020), in view of the good effectiveness of cART in Shenzhen; a uniform (0, 50) for the origin of the epidemic, considering the first reported HIV-positive case of mainland China in 1985; a log-normal (0, 1) for the reproduction number, with a 95 per cent interval from 0.14 to 7.1; and a Beta (1, 10) for the sampling proportion, in view of the calculated sampling rate in Supplementary Table S3. The time to most recent common ancestor (tMRCA) and evolutionary rate were estimated. The Monte Carlo Markov Chains (MCMC) were run for at least 500 million generations until effective sample size (ESS) of all parameters >200 after an initial 10 per cent as burn-in in Tracer v1.7.1 (Rambaut et al. 2018). The trajectories of key epidemiological parameters through time were plotted using the phylodynamics script (https://github.com/BEAST2-Dev/phylodynamics) in R version 4.0.2.

2.4 Reconstruction of viral diffusion history among local residents and immigrants

To understand the spatial diffusion history of each lineage among populations, the phylogeographic analysis was performed to reconstruct the viral migrations under GTR + I + G4 nucleotide

substitution model with the default distribution priors, the relaxed log-normal molecular clock model and the Bayesian Sky-Grid model in BEAST v1.10.4 (Gill et al. 2013; Graf et al. 2015; Vrancken et al. 2020). The household register locations were used as discrete traits and the transition rates were evaluated using a discrete asymmetric substitution model by means of Bayesian stochastic search variable selection (Faria et al. 2019), of which the support for viral migration between populations was calculated using Bayes Factor (BF) in SpreaD3 (Bielejec et al. 2016). The MCMC analyses were run for enough generations to ensure the high ESS of >200 after an initial 10 per cent as burn-in in Tracer v1.7.1 (Rambaut et al. 2018). The fractions of all strongly supported transmission events (BF > 10) were drawn using Sankey plot in R.

Importantly, in order to reduce the sampling biases at the number of sequences with different household register locations, we randomly down-sampled sequences by household register location for three times: for L1, L4, and L5, with the maximum number of 20 sequences; for L2, with the maximum number of 10 sequences (Supplementary Table S4) (Graf et al. 2015).

2.5 Statistical analyses

Sex, age, risk group, and household among different HIV-1 CRF01_AE lineages were compared using chi-square test and Fisher's exact test in SPSS software v 26. *P*-values <0.05 denoted statistically significant differences.

2.6 Sequence accession numbers

The sequences sampled before 2013 had been submitted to GenBank previously (Zhao et al. 2016). The remaining sequences were submitted to GenBank (the accession numbers: MW270172–MW270928).

3. Results

3.1 The sociodemographic characteristics of CRF01_AE lineages in Shenzhen

Of the 1,432 individuals, 86.8 per cent were males, 55 per cent were MSM, 37.3 per cent were heterosexuals, and 6.3 per cent were PWID (Table 1). A total of 1,229 (86 per cent) were immigrants, originating from 31 of 34 province/regions in China and four countries (Thailand, Vietnam, Myanmar, and the Netherlands). The demographic characteristics of individuals infected by CRF01_AE were consistent with those of the overall HIV-1-infected population in Shenzhen during this period (Zhao et al. 2016). Eleven distinct CRF01_AE lineages containing 1,377 sequences were identified, denoted SZ-L1-11 (Fig. 1 and Table 1). Of the eleven lineages, five belong to the previously reported lineages in the national study (Feng et al. 2013), so we named them using their original order, and the other six lineages were ordered by the number of sequences in the lineages. The ranges of intra-lineage genetic distances were 0.7–3.4 per cent (Supplementary Table S1). L4 and 5 accounted for >70 per cent of sequences (1,006/1,432).

A comparison of the sociodemographic information among lineages ($n \geq 10$) was performed. Significant differences ($P < 0.05$) in the variables of sex, age, risk group, and household register were found (Table 1). Obviously, males with L4–L6 accounted for >98 per cent of cases. The individuals born after 1990 with L4–L6 was >13 per cent, while the individuals born before 1970 with L1–L3 and L7 was >20 per cent. Heterosexuals and PWID accounted for >90 per cent in L1–L3 and L7–L8, while L4–L6 were more common among MSM (range: 61.1–78.2 per cent). Notably, the proportion of males and females among heterosexuals with L4–L6 lineages were significantly unbalanced (22.6 per cent vs. 0.8 per cent in L4, 17.2 per cent vs. 1.4 per cent in L5, and 38.9 per cent vs. 0 per cent in L6), compared with that with L1–L3 and L7–L8 lineages (38.2 per cent vs. 35.1 per cent in L1, 33.3 per cent vs. 42 per cent in L2, 16.7 per cent vs. 41.7 per cent in L3, 45.5 per cent vs. 36.4 per cent in L7, and 40 per cent vs. 40 per cent in L8).

Significant differences of the sociodemographic information were also observed between individuals within lineages and those without (Table 1). Younger individuals born after 1990, males, and MSM were more likely to be in lineages. The immigrants were more likely to have HIV-1 CRF01_AE lineages than local residents in Shenzhen.

3.2 The transmission dynamics of CRF01_AE lineages

Four lineages, containing ≥20 sequences, had good temporal signals evaluated in TempEst (Wilkinson et al. 2019) (Supplementary Figure S1) and were selected for further time-calibrated reconstruction (Vasylyeva et al. 2019). Molecular clocks revealed that their median tMRCA ranged between 1992 and 1999, and the median evolutionary rates were 1.85–3.26 × 10^{-3} substitutions/site/year. L2 appeared around 1992, followed by L1, L4, and L5 at the end of 2000s (Table 2).

Despite sharing same priors, the median basic viral reproduction numbers (R_0) of four lineages differed significantly, ranging 1.71–3.92; of those, L1 was highest close to 4, followed by L2, L4, and L5 (Table 2). Notably, the noninfectious rates for L1 and L2 were much lower than those for the other two lineages, indicating that they had longer infectious periods and more chances to infect the susceptible individuals (Table 2). Following a stable period, the effective viral reproduction numbers (R_e) of all lineages began to decline, but the time of plateau and the decrease in rate differed across lineages (Fig. 2). Although the initial decline of R_e for L1 and L2 occurred before 2005, the R_e of L1 had fell <1 before 2010 and the R_e of L2 declined to 1 about 5 years later than L1. The temporal changes in R_e for L4 and L5 were similar, showing an initial decline during 2005–2010, and a value close to 1 at the end of sampling. As shown in SIR trajectory, the estimated numbers of the initial susceptible individuals in four lineages were significantly different, and L1 had nearly achieved a complete depletion at the end of sampling. Moreover, according to the curves of incidence and prevalence, L1 has reached the epidemic peak followed by a rapid decline, while L2 had just passed the peak and L4 and L5 were still before the peak. Importantly, the prevalent curve should represent the level of the patients with the high HIV transmission risk, who are not under cART after infection or do not achieve virological suppression, rather than the whole HIV infection population.

3.3 The viral migration of CRF01_AE lineages among local residents and immigrants

The phylogeographic results, combining three random down-sampled datasets, revealed the complex and distinct viral migration history for lineages. In L1 and L2 that were prevalent among heterosexuals and PWID, >90 per cent of the viral migration events sourced from immigrants from other cities in Guangdong province and the local residents of Shenzhen (BF > 10) (Fig. 3A and B). For two lineages among MSM, the viral spread from the immigrants of Jiangxi and Shaanxi provinces accounted for 57.5 per

Table 1. The sociodemographic characteristics of individuals included in or not in lineages.

	Total	SZ-L1	SZ-L2	SZ-L3	SZ-L4	SZ-L5	SZ-L6	SZ-L7	SZ-L8	SZ-L9	SZ-L1	SZ-L1	Chi-square/ Fisher's exact test[a]	Not in lineage	Chi-square/ Fisher's exact test[b]
Categories	N (%)	N (%)	N (%)	N (%)	N (%)	N (%)	N (%)	N (%)	N (%)	N (%)	N (%)	N (%)	P-value	N (%)	P-value
Sex															
Male	1,243 (86.8)	135 (59.2)	43 (53.1)	7 (58.3)	714 (99)	281 (98.6)	18 (100)	7 (63.6)	4 (40)	1 (25)	2 (50)	2 (66.7)	<0.001	29 (52.7)	<0.001
Female	189 (13.2)	93 (40.8)	38 (46.9)	5 (41.7)	7 (1)	4 (1.4)	–	4 (36.4)	6 (60)	3 (75)	2 (50)	1 (33.3)		26 (47.3)	
Born period															
Before 1960	54 (3.8)	17 (7.5)	7 (8.7)	–	7 (1.0)	8 (2.8)	–	5 (45.5)	–	–	1 (25)	–	<0.001	9 (16.3)	0.001
1960s	116 (8.1)	32 (14.0)	10 (12.3)	3 (25.0)	34 (4.8)	29 (10.2)	2 (11.1)	1 (9.1)	1 (10)	1 (25)	–	–		3 (5.5)	
1970s	374 (26.1)	76 (33.3)	23 (28.4)	2 (16.7)	172 (23.8)	74 (26.0)	6 (33.3)	2 (18.2)	4 (40)	–	1 (25)	2 (66.7)		12 (21.8)	
1980s	695 (48.5)	90 (39.5)	37 (45.7)	6 (50.0)	376 (52.1)	137 (48.1)	5 (27.8)	3 (27.3)	4 (40)	3 (75)	2 (50)	1 (33.3)		31 (56.4)	
After 1990	193 (13.5)	13 (5.7)	4 (4.9)	1 (8.3)	132 (18.2)	37 (13.0)	5 (27.8)	–	1 (10)	–	–	–		–	
Risk[c]															
MSM	787 (55)	7 (3.1)	1 (1.2)	1 (8.3)	529 (73.4)	223 (78.2)	11 (61.1)	1 (9.1)	1 (10)	–	–	1 (33.3)	<0.001	12 (21.8)	<0.001
HET(M)	366 (25.6)	87 (38.2)	27 (33.3)	2 (16.7)	163 (22.6)	49 (17.2)	7 (38.9)	5 (45.5)	4 (40)	1 (25)	2 (50)	1 (33.3)		18 (32.7)	
HET(F)	167 (11.7)	80 (35.1)	34 (42)	5 (41.7)	6 (0.8)	4 (1.4)	–	4 (36.4)	4 (40)	3 (75)	2 (50)	1 (33.3)		24 (43.6)	
PWID(M)	82 (5.7)	43 (18.9)	13 (16)	4 (33.3)	13 (1.8)	8 (2.8)	–	1 (9.1)	–	–	–	–		–	
PWID(F)	9 (0.6)	4 (1.8)	3 (3.7)	–	1 (0.1)	–	–	–	–	–	–	–		1 (1.8)	
Unknown/Others	21 (1.5)	7 (3.1)	3 (3.7)	–	9 (1.2)	1 (0.4)	–	–	1 (10)	–	–	–		–	
Household															
Shenzhen	203 (14.2)	18 (7.9)	8 (9.9)	2 (16.7)	104 (14.4)	41 (14.4)	3 (16.7)	5 (45.5)	3 (30)	–	2 (50)	–	<0.001	17 (30.9)	0.008
Southern (except SZ)	383 (26.7)	115 (50.4)	44 (54.3)	4 (33.3)	161 (22.3)	42 (14.7)	5 (27.8)	–	1 (10)	–	–	–		11 (20)	
Central	354 (24.7)	43 (18.9)	14 (17.3)	1 (8.3)	182 (25.2)	88 (30.9)	6 (33.3)	2 (18.2)	3 (30)	3 (75)	–	–		12 (21.8)	
Southwestern	199 (13.9)	36 (15.8)	11 (13.6)	1 (8.3)	101 (14)	34 (11.9)	3 (16.7)	3 (27.3)	3 (30)	–	1 (25)	1 (33.3)		5 (9.1)	
Eastern	151 (10.5)	8 (3.5)	1 (1.2)	3 (25)	94 (13)	38 (13.3)	1 (5.6)	1 (9.1)	–	–	–	–		5 (9.1)	
Northwestern	48 (3.4)	4 (1.8)	–	0	33 (4.6)	9 (3.2)	–	–	–	–	–	–		2 (3.6)	
Northeastern	57 (4)	2 (0.9)	–	1 (8.3)	29 (4)	24 (8.4)	–	–	–	–	–	1 (33.3)		–	
Northern	23 (1.6)	–	–	–	14 (2.1)	8 (2.8)	–	–	–	–	–	–		1 (1.8)	
Hong Kong, Macao, Taiwan	8 (0.6)	2 (0.9)	1 (1.2)	–	2 (0.3)	–	–	–	–	1 (25)	1 (25)	1 (33.3)		–	
Outside China	6 (0.4)	–	2 (2.4)	–	1 (0.1)	1 (0.4)	–	–	–	–	–	–		2 (3.6)	
Total	1,432 (100)	228 (100)	81 (100)	12 (100)	721 (100)	285 (100)	18 (100)	11 (100)	10 (100)	4 (100)	4 (100)	3 (100)	–	55 (100)	–

[a]Chi-square/Fisher's exact test among lineages.
[b]Chi-square/Fisher's exact test in and out lineages.
[c]MSM: men who have sex with men; HET(M): male heterosexuals; HET(F): female heterosexuals; PWID(M): male people who inject drugs; PWID(F): female people who inject drugs, which was self-reported during investigation.

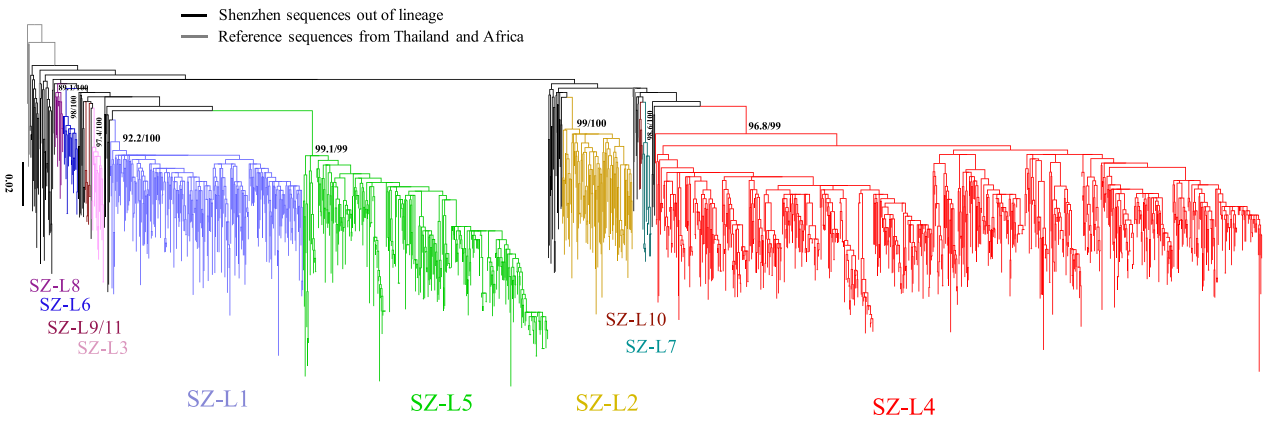

Figure 1. The maximum likelihood phylogenetic tree of HIV-1 CRF01_AE pol sequences in Shenzhen.

The colored branch represents CRF01_AE lineages in Shenzhen, and the black branches were reference sequences used to distinguish the lineages and fix the topology of tree.

Table 2. The phylodynamic estimates for four CRF01_AE lineages in Shenzhen under SIR model in BEAST v 2.6.2.

	SZ-L1 (95% HPD)	SZ-L2 (95% HPD)	SZ-L4 (95% HPD)	SZ-L5 (95% HPD)
tMRCA	1998.10 (1995.43–2000.24)	1992.53 (1985.11–1998.00)	1998.25 (1996.13–2000.12)	1999.15 (1996.29–2001.62)
Evolutionary rate (x10^{-3}, s/s/y)	2.08 (1.62–2.61)	1.85 (1.20–2.62)	3.26 (2.85–3.69)	2.17 (1.81–2.54)
Basic reproduction number, R_0	3.92 (2.24–6.29)	2.65 (1.31–5.21)	1.96 (1.25–3.20)	1.71 (1.17–2.62)
Noninfectious rate, δ	0.36 (0.18–0.60)	0.31 (0.08–0.81)	0.51 (0.12–1.16)	0.66 (0.16–1.63)
Duration of infection, D (years)[a]	2.78 (1.67–5.56)	3.22 (1.23–12.5)	1.96 (0.86–8.33)	1.52 (0.61–6.25)
Sampling proportion, p (x10^{-3})	3.4 (0.4–10)	1.6 (1–8.9)	34.9 (2.35–118.3)	29.2 (0.88–106.2)

[a]Calculated with the equation: $D = 1/\delta$. HPD: The highest posterior density interval.

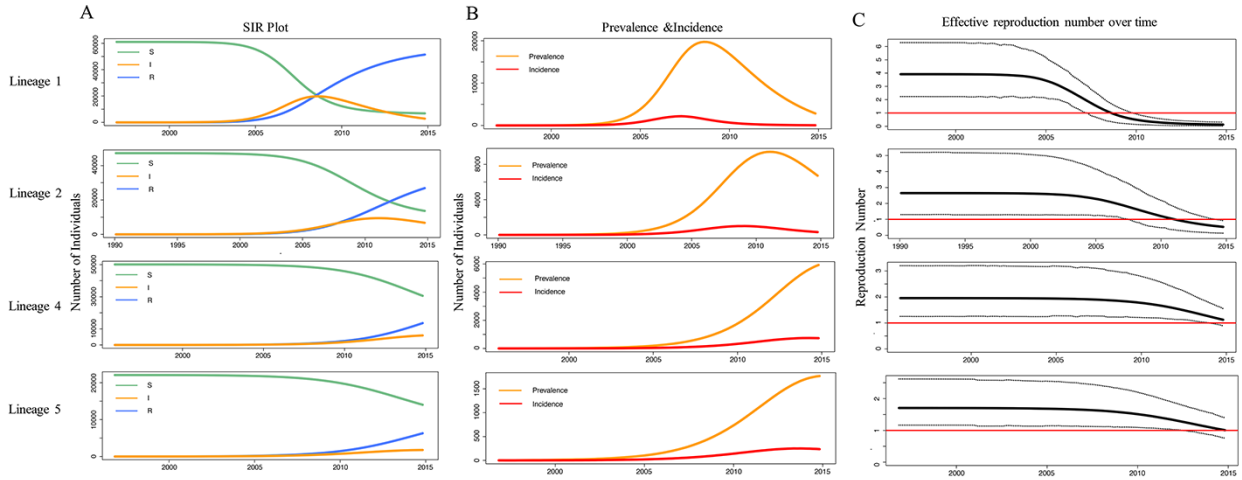

Figure 2. The reconstructed SIR trajectory, incidence, and effective reproduction number of four CRF01_AE lineages over time.

The epidemic dynamics of four lineages were plotted using the phylodynamics script in R. (A) The plot of overall SIR dynamics over time; (B) the plot of the prevalence and incidence over time; (C) the plot of the estimated effective reproduction number (R_e) over time. The black dotted lines represent the 95 per cent HPD interval and the red dotted line represents epidemiological threshold ($R_e = 1$).

cent and 35.8 per cent of all migration events in L4 (Fig. 3C), while the immigrants from Hubei province were the dominant sources (68.1 per cent) in L5 (Fig. 3D). However, for all four lineages, no predominant population was observed as major sinks or hubs of viral diffusion (Fig. 3).

4. Discussion

Shenzhen, as China's first special economic zone, attracts a large number of migration workers from other regions across China since the 1980s, most of who would return to their home province after some time or during Chinese traditional holidays

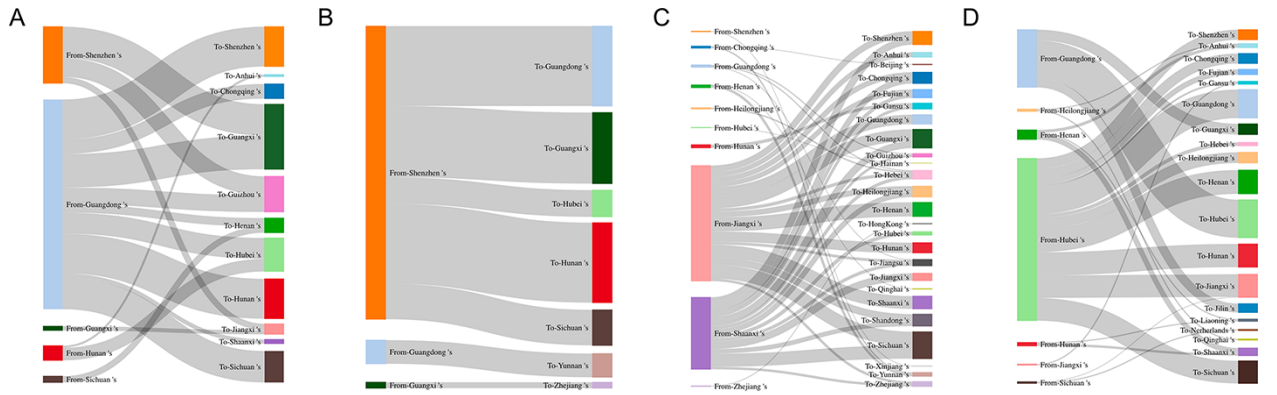

Figure 3. The viral migration events among population of four major CRF01_AE lineages in Shenzhen.

The links and directions from sources toward recipients (BF > 10) are drawn using Sankey plot in R. The left and right sides of Sankey plot represent the source and recipient, respectively. The fractions of the two sides represent the proportions of migration events from sources toward recipients. SZ-L1(A); SZ-L2(B); SZ-L4(C); SZ-L5(D).

(Yang et al. 2020). Therefore, Shenzhen could serve as a hub for HIV-1 dissemination and may become the epitome and predictor of the national HIV epidemic.

We found eleven CRF01_AE lineages with high branch support value. Of them, five lineages belong to the previously reported top five CRF01_AE lineages throughout China (Feng et al. 2013; Li et al. 2017; Wang et al. 2017): L1–L3 were prevalent among heterosexuals/PWID and L4–L5 were common among MSM. Moreover, they accounted for 93 per cent of all CRF01_AE cases, indicating that the HIV-1 epidemic in Shenzhen is representative of the national HIV epidemic. Of them, four main lineages had distinct origin times (Table 2), over the 1990s, and all estimated tMRCAs were close to or later than those in the national reports (Feng et al. 2013), suggesting that although Shenzhen owned more migrants historically, it was not the origin that drove CRF01_AE spread across China and more likely to be the cross-center of HIV epidemic. Of eleven HIV-1 CRF01_AE lineages circulating in Shenzhen, although the remaining six lineages accounted for 3 per cent, they also need to be concerned about monitoring their further developments.

It is well estimated that sexual contacts were the predominant high-risk behavior for CRF01_AE infection in Shenzhen (Zhao et al. 2016). Obviously, the ratio of males and females was inconsistent among heterosexuals in lineages L4–L6 that were prevalent among MSM; thus, further investigation was performed to re-evaluate the risk behavior of 219 male heterosexuals in L4–L6. We found that twenty-eight individuals admitted to be MSM, while others were unwilling to disclose this information to the investigators or migrated to other regions. These findings suggested that some heterosexual males might have concealed their homosexual orientation initially (Shang et al. 2012), which may lead an underestimation of the HIV-1 prevalence among MSM.

R_0 is an indicator of the number of new infections caused by one positive individual during his infectious life in a population where everyone is susceptible. The changes in this number over time, termed R_e, can further reflect the viral transmission dynamics under interventions. The R_0 differed for four lineages, but their R_e had declined over time (Fig. 3). For L1 and L2, the obvious decline of R_e coincided with the initiation of the needle exchange programs from 2000, which freely provided needles (Liu, Sullivan, and Wu 2007), the cART around 2003 (Zhang et al. 2007; Cao, Hsieh, and Li 2020), and the wide sex education during the 2000s (Rou et al. 2010). During the initial period of HIV-1 epidemic, >40 per cent of infections were in PWID both in Shenzhen and nationwide (Peng 2004; Qian et al. 2006). This may explain why the R_0 of L1 and L2 were very high, while the R_e rapidly declined after the needle exchange programs (Liu, Sullivan, and Wu 2007).

Differently, for L4 and L5, the obvious decline in R_e did not appear until 2010, suggesting that the routine HIV interventions would not effectively control the viral transmission among MSM. This was attributed to the unique characteristics of MSM, which may complicate their detection and treatment before the initiation of universal test and treat strategies (performed after 2016). Since 2008, a large-scale MSM surveillance program was launched in Shenzhen, identifying a large number of MSM who would be encouraged to detect HIV, possibly resulting in the rapid reduction of R_e in L4 and L5 after 2010. The traditional epidemiological investigation performed by Shenzhen CDC in 2014 also proved that the prevalence of HIV among MSM had started to decline along the surveillance program of MSM. Thus, we are confident that the wide-surveillance program, which led to more diagnoses, was effective in controlling the HIV epidemic among MSM. The various transmission dynamics indicated that conventional strategies did not equally control various lineages, possibly resulting from their acceptance degree to treatment and the coverage range of interventions for risk populations within lineages. The inconsistences of epidemic dynamics among heterosexuals/PWID and MSM suggest that the further targeting interventions, such as behavior and health education among the active MSM or enhanced managements of detection and treatment, would be more important.

The key of erasing HIV epidemic is to characterize who will infect others most possibly or be infected at the most risk (Abeler-Dorner et al. 2019). Reconstruction of the viral migration history of CRF01_AE lineages among population will help identify the transmission source, sink, and hub. For L1 and L2, the higher fraction of viral migration events occurred from the immigrants of other cities in Guangdong province or Shenzhen local residents toward the immigrants of other provinces (Fig. 3A and 3B), implying the residents of Guangdong province were the main source of the CRF01_AE dispersal among heterosexuals and PWID. Although these two lineages have been controlled, the continued HIV intervention programs are still important to enhance and solidify the effects of HIV prevention among heterosexuals and PWID. MSM accounted for 55 per cent of all cases (Table 1), so two lineages (L4 and L5) that had yet been controlled

completely became the urgent targets to reduce HIV transmission. It is noted that the immigrants from Jiangxi, Shaanxi, and Hubei provinces were the main sources of viral spread among MSM. The Shenzhen statistical report (2015) showed that the immigrants in Shenzhen mainly came from Guangdong (35 per cent), followed by Hunan, Sichuan, and Guangxi, while the immigrants from Hubei, Jiangxi, and Shaanxi provinces accounted for 4 per cent, 4 per cent, and 1 per cent. Although the three provinces were not the top import provinces, the immigrants from them can drive CRF01_AE spread at large scale among MSM. No obvious sinks for all four lineages suggest that the mixed transmissions among populations occurred since the virus was brought to Shenzhen by immigrants and the interventions targeting transmission sources can get the maximum effects on the prevention and control of HIV.

There are some limitations that might impact phylogenetic reconstructions: (1) most sequences were from newly diagnosed cases and due to the development of detection ability, the sampling depth might vary throughout time. (2) The overrepresentation of sequences in the high-density sampling populations with different household register locations might bias the phylogeographic reconstruction and we performed the down-sampling to solve it. (3) The update of cART policies over time in China would shorten the duration of infectiousness, one dimension setting for the becoming-non-infectious rate in phylodynamic analysis might underestimate the effectiveness of interventions; one assumption of SIR model is that the population size is constant, and Shenzhen is a city of immigrants with a growing population along with the frequent population exchange, which would bias our phylodynamic reconstruction using SIR and overestimate the effectiveness of interventions. Therefore, in spite of the declining epidemic of all four main CRF01_AE lineages estimated in this study, the continued public health interventions should be taken to keep and strengthen the controlled effects. Finally, HIV/AIDS healthcare in Shenzhen begun early in the early 2000 and covered both the local residents and the immigrants, which would not impact the estimates of transmission history among populations.

In conclusion, multiple independent introductions shaped the HIV-1 CRF01_AE epidemics in Shenzhen, most of which were drove by immigrants. Although routine interventions can effectively decrease the prevalence of CRF01_AE, the epidemic patterns of the lineages were variable. These results suggest that more precisely targeting interventions, focusing on certain transmission sources, would exert a maximum effect on controlling the whole HIV-1 epidemic in Shenzhen and the continued molecular tracing would improve HIV-1 comprehensive prevention and control.

Supplementary data

Supplementary data is available at *Virus Evolution* online.

Acknowledgements

The authors would like to thank the staff of the First Affiliated Hospital of China Medical University and Shenzhen Center for Disease Control and Prevention for recruiting the participants. They also would like to thank all the participants of this study.

Funding

Mega-projects of National Science Research for the 13th Five-Year Plan (2017ZX10201101, 2018ZX10721102-006-003); National Natural Science Foundation of China (82072272, 81573211, 81270043); National Key S&T Special Projects on Major Infectious Diseases (2018ZX10721102); Guangdong Basic and Applied Basic Research Foundation (2019B1515120003); San-Ming Project of Medicine in Shenzhen (SZSM201811071); Shenzhen Key Medical Discipline Construction Fund (SZXK064).

Conflict of interest: All authors declared no conflicts of interest.

Author contributions

H.S., J.Z., and X.H. designed the study; C.Z., H.L., L.C., Z.Y., Y.G., and J.Z. collected the blood samples and demographic information, and amplified viral sequences; M.A. analyzed the data and wrote the primary draft; X.H. and J.Z. revised the manuscript; the final version of this manuscript was approved by all authors.

References

Abeler-Dorner, L. et al. (2019) 'Pangea-hiv 2: Phylogenetics and Networks for Generalised Epidemics in Africa', *Current Opinion in Hiv and Aids*, 14: 173–80.

Abubakar, Y. F. et al. (2013) 'Multiple Independent Introductions of Hiv-1 Crf01_ae Identified in China: What are the Implications for Prevention?' *PLoS One*, 8: e80487.

An, 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', *Journal of Virology*, 86: 12402–6.

Angelis, K. et al. (2015) 'Global Dispersal Pattern of Hiv Type 1 Subtype Crf01_ae: A Genetic Trace of Human Mobility Related to Heterosexual Sexual Activities Centralized in Southeast Asia', *The Journal of Infectious Diseases*, 211: 1735–44.

Bielejec, F. et al. (2016) 'Spread3: Interactive Visualization of Spatiotemporal History and Trait Evolutionary Processes', *Molecular Biology and Evolution*, 33: 2167–9.

Bui, Q. M. et al. (posting date). Iq-tree version 1.6.8: Tutorials and manual phylogenomic software by maximum likelihood.

Cao, W., Hsieh, E., and Li, T. (2020) 'Optimizing Treatment for Adults with Hiv/aids in China: Successes over Two Decades and Remaining Challenges', *Current HIV/AIDS Reports*, 17: 26–34.

Cassels, S. (2020) 'Time, Population Mobility, and Hiv Transmission', *Lancet HIV*, 7: e151–2.

Cheng, H. et al. (1994) 'Hiv-1 Subtype E in Yunnan, China', *Lancet*, 344: 953–4.

Dennis, A. M. et al. (2020) 'Human Immunodeficiency Virus Type 1 Phylodynamics to Detect and Characterize Active Transmission Clusters in North Carolina', *The Journal of Infectious Diseases*, 221: 1321–30.

Faria, N. R. et al. (2019) 'Distinct Rates and Patterns of Spread of the Major Hiv-1 Subtypes in Central and East Africa', *PLoS Pathogens*, 15: e1007976.

Feng, Y. 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: 1793–802.

Gill, M. S. et al. (2013) 'Improving Bayesian Population Dynamics Inference: A Coalescent-based Model for Multiple Loci', *Molecular Biology and Evolution*, 30: 713–24.

Graf, T. et al. (2015) 'Contribution of Epidemiological Predictors in Unraveling the Phylogeographic History of Hiv-1 Subtype C in Brazil', *Journal of Virology*, 89: 12341–8.

He, Y. et al. (2015) Effects of Anti-hiv Therapy on Aids Patients in Shenzhen. The 7th National conference on AIDS and hepatitis C of Chinese Medical Association, Shenzhen, China, pp. 190–1.

Jia, D. et al. (2019) 'Two-year Cross-sectional Studies Reveal that Single, Young Msms in Shenzhen, China are at High Risk for Hiv Infection', *Virology Journal*, 16: 83.

Kostaki, E. G. et al. CoRis. (2019) 'Spatiotemporal Characteristics of the Largest Hiv-1 Crf02_ag Outbreak in Spain: Evidence for Onward Transmissions', *Frontiers in Microbiology*, 10: 370.

Kuhnert, D. et al. (2014) 'Simultaneous Reconstruction of Evolutionary History and Epidemiological Dynamics from Viral Sequences with the Birth-death Sir Model', *Journal Of The Royal Society, Interface / the Royal Society*, 11: 20131106.

LANL. (2019) posting date. Hiv sequence database: Hiv circulating recombinant forms (crfs).

Li, X. et al. (2017) 'Tracing the Epidemic History of Hiv-1 Crf01_ae Clusters Using Near-complete Genome Sequences', *Scientific Reports*, 7: 4024.

Liu, B., Sullivan, S. G., and Wu, Z. (2007) 'An Evaluation of Needle Exchange Programmes in China', *AIDS*, 21: S123–8.

Nguyen, L. T. et al. (2015) 'Iq-tree: A Fast and Effective Stochastic Algorithm for Estimating Maximum-likelihood Phylogenies', *Molecular Biology and Evolution*, 32: 268–74.

Peng, C. (2004) 'Aids Epidemic Status in Shenzhen, 1992–2003', *South China J Pre Med*, 30: 20,22.

Peng, X. et al. (2015) 'Heterogeneous Evolution of Hiv-1 Crf01_ae in Men Who Have Sex with Men (Msm) and Other Populations in China', *PLoS One*, 10: e0143699.

Qian, H. Z. et al. (2006) 'Injection Drug Use and Hiv/aids in China: Review of Current Situation, Prevention and Policy Implications', *Harm Reduction Journal*, 3: 4.

Rambaut, A. et al. (2018) 'Posterior Summarization in Bayesian Phylogenetics Using Tracer 1.7', *Systematic Biology*, 67: 901–4.

——— et al. (2016) 'Exploring the Temporal Structure of Heterochronous Sequences Using Tempest (Formerly Path-o-gen)', *Virus Evolution*, 2: vew007.

Ratmann, O. et al. Rakai Health Sciences P, the Pangea HIVC. (2020) 'Quantifying Hiv Transmission Flow between High-prevalence Hotspots and Surrounding Communities: A Population-based Study in Rakai, Uganda', *Lancet HIV*, 7: e173–8.

Rou, K. et al. (2010) 'Scaling up Prevention Programmes to Reduce the Sexual Transmission of Hiv in China', *International Journal of Epidemiology*, 39: ii38–46.

Shang, H. et al. (2012) 'Hiv Prevention: Bring Safe Sex to China', *Nature*, 485: 576–7.

Shenzhen-Statistics. (2020) posting date. The seventh national population census of the people's republic of China. Shenzhen Statistics Bureau.

Tan, W. et al. (2016) 'Epidemiological Characteristics of Hiv/aids in Shenzhen City, 2013–2015', *China Tropical Medicine*, 16: 782–784,788.

UNAIDS. (2020) posting date. Unaids Data 2020.

Vasylyeva, T. I. et al. (2019) 'Tracing the Impact of Public Health Interventions on Hiv-1 Transmission in Portugal Using Molecular Epidemiology', *The Journal of Infectious Diseases*, 220: 233–43.

Vrancken, B. et al. (2020) 'Comparative Circulation Dynamics of the Five Main Hiv Types in China', *Journal of Virology*, 94: e00683–20.

Wang, N., and Zhong, P. (2015) 'Molecular Epidemiology of Hiv in China: 1985–2015', *Chinese Journal of Epidemiology*, 36: 541–6.

Wang, X. et al. (2017) 'Phylodynamics of Major Crf01_ae Epidemic Clusters Circulating in Mainland of China', *Scientific Reports*, 7: 6330.

Wilkinson, E. et al. (2019) 'The Effect of Interventions on the Transmission and Spread of Hiv in South Africa: A Phylodynamic Analysis', *Scientific Reports*, 9: 2640.

Yang, S. et al. (2018) 'Hiv Epidemics in Shenzhen and Chongqing, China', *PLoS One*, 13: e0192849.

Yang, Z. et al. (2020) 'Spatiotemporal Patterns of Population Mobility and Its Determinants in Chinese Cities Based on Travel Big Data', *Sustainability*, 12: 4012.

Zhang, F. et al. (2011) 'Effect of Earlier Initiation of Antiretroviral Treatment and Increased Treatment Coverage on Hiv-related Mortality in China: A National Observational Cohort Study', *The Lancet Infectious Diseases*, 11: 516–24.

——— et al. (2007) 'The Chinese Free Antiretroviral Treatment Program: Challenges and Responses', *AIDS*, 21: S143–8.

Zhao, G. et al. (2012) 'Study on the Molecular-epidemiological Characteristics of Hiv-1 in Shenzhen, 1992–2008', *Chinese Journal of Epidemiology*, 33: 82–7.

Zhao, J. et al. (2016) 'The Dynamics of the Hiv Epidemic among Men Who Have Sex with Men (Msm) from 2005 to 2012 in Shenzhen, China', *Scientific Reports*, 6: 28703.