

## RESEARCH ARTICLE

# Complete genome sequencing and evolutionary analysis of HCV subtype 6xg from IDUs in Yunnan, China

Min Chen<sup>1</sup>, Yanling Ma<sup>1</sup>, Huichao Chen, Jie Dai, Hongbing Luo, Manhong Jia\*, Zhizhong Song\*

Institute for AIDS/STD Control and Prevention, Yunnan Center for Disease Control and Prevention, Kunming, Yunnan, China

☯ These authors contributed equally to this work.

\* [songzhizhong@yncdc.cn](mailto:songzhizhong@yncdc.cn) (ZS); [jiamanhong@hotmail.com](mailto:jiamanhong@hotmail.com) (MJ)



## OPEN ACCESS

**Citation:** Chen M, Ma Y, Chen H, Dai J, Luo H, Jia M, et al. (2019) Complete genome sequencing and evolutionary analysis of HCV subtype 6xg from IDUs in Yunnan, China. PLoS ONE 14(5): e0217010. <https://doi.org/10.1371/journal.pone.0217010>

**Editor:** Chiyu Zhang, Institut Pasteur of Shanghai Chinese Academy of Sciences, CHINA

**Received:** November 16, 2018

**Accepted:** May 2, 2019

**Published:** May 16, 2019

**Copyright:** © 2019 Chen et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** The nucleotide sequences obtained in this study were submitted to GenBank under the accession numbers MK139015-MK139022.

**Funding:** This work was supported by the National Natural Science Foundation of China (81560327). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

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

## Abstract

### Background

HCV genotype 6 (HCV-6) typically circulates in Southeast Asia and exhibits the highest genetic diversity among the eight HCV genotypes. In our previous work, a group of HCV-6 sequences was not clearly classified. Here, we further characterized this HCV-6 variant and analyzed the evolutionary history of the enlarged HCV-6 family.

### Methods

Blood samples from eight HCV seropositive samples collected from intravenous drug users (IDUs) in 2014 in Yunnan Province, China. The full-length HCV genome sequences were amplified by using reverse transcription PCR followed by DNA sequencing and phylogenetic analysis. Bayesian evolutionary analysis was performed with the complete coding region sequences of subtype 6a-6xh.

### Results

The eight genomes had the same coding region of 9051 nucleotides. The complete coding region sequences of the eight HCV isolates formed a distinct phylogenetic group from the previously assigned HCV-6 subtypes (6a-6xf), however which clustered with 6xg reference sequences that were found in Kachin State, Myanmar, and recently assigned and released. The p-distances of the eight isolates to subtype 6a-6xf and 6xh ranged from 0.143 to 0.283. Based on the HCV-6 complete coding region sequences, we constructed a timescaled phylogenetic tree to reveal the HCV-6 evolutionary history, in which there were four HCV-6 phylogenetic subsets, whose median tMRCAs were 294.8, 388.5, 348.5 and 197.0 years ago, respectively. Subtype 6xg clustered into Subset I, and had the most recent common ancestor with subtype 6n, which dated back to 101.2 (95% HPD: 78.7, 125.8) years ago. The genetic evolutionary analysis further confirmed that subtype 6xg originated from Myanmar, and transmitted to Dehong through cross-border IDUs.

## Conclusion

The HCV-6 variant characterized in this study belonged to newly assigned subtype 6xg. Our finding further confirmed the assignment of 6xg. HCV-6 family was highly diverse and had a complicated evolutionary history in Southeast Asia. It is necessary to further characterize HCV-6 genetics in this region.

## Introduction

Hepatitis C virus (HCV) belongs to the genus *Hepacivirus* in the family *Flaviviridae*. It has a single-stranded positive RNA genome of approximately 9600 nucleotides in length. The genome contains a single open reading frame (ORF) encoding a polyprotein, which further produces three structural (Core, E1, and E2) and seven non-structural proteins (p7, NS2, NS3, NS4A, NS4B, NS5A and NS5B). The untranslated regions (UTRs) flank the ORF in 5' and 3' terminal, which involves in protein translation and RNA replication. As a blood-borne pathogen, HCV infects an estimated 143 million people (2%) worldwide as of 2015 [1]. Persistent HCV infection is one of major causes of liver cirrhosis, hepatocellular carcinoma and liver failure [2].

The first full-length genome sequence of HCV was reported in 1989 [3], which promoted the development of HCV diagnosis and antiviral treatment [4]. However, HCV exhibits a high genetic diversity, the isolates from the different areas showed heterogeneity at the nucleotide level. Furthermore, these different isolates showed distinct origins, epidemiological history, even response to antiviral treatment [5, 6]. Thus, the classification of HCV is an important issue in HCV research. HCV classification is based on phylogenetic analyses of genome sequences, and include the genotype and subtype levels [7, 8]. Nowadays, HCV classified into eight genotypes, each of which contains a different number of subtypes. As of 2017, the number of confirmed subtypes increased to 86. Among the eight genotypes, genotype 6 predominantly distributes in South East Asia, where the most of subtypes from genotype 6 were identified [9].

Yunnan, a province in southwest China, borders Myanmar, Laos and Vietnam. Historically, Yunnan was a channel of heroin into China. Drug trafficking and intravenous drug use drove the development of HIV epidemic in Yunnan [10]. Stemmed from Yunnan, the HIV epidemic further spread to the other parts of China [11, 12]. Because intravenous drug use is also a major risk factor for HCV infection, co-infection with HCV and HIV-1 were common among IDUs in Yunnan. Recently, we carried out an HCV molecular epidemiological study among HIV-infected IDUs in Yunnan, which revealed the complicated and distinctive HCV genetics [13]. However, there was a group of HCV-6 sequences that were not clearly classified. In this study, we sequenced and characterized the full-length genome sequences of this group of isolates.

## Materials and methods

### Subjects and specimens

A total of eight HCV seropositive samples were collected from IDUs in Dehong Prefecture, China. The study subjects provided written informed consent. The study was approved by Biomedical Ethics Review Committee of Yunnan Center for Disease Control and Prevention.

## PCR amplification and sequencing

Plasma was separated from whole blood. Viral RNA was extracted from 140  $\mu$ l of plasma by using the QIAamp Viral RNA Mini kit (Qiagen, Valencia, CA, United States) according to the manufacturer's instructions. The cDNA was synthesized by using PrimeScript<sup>TM</sup> 1<sup>st</sup> Strand cDNA Synthesis Kit (Takara, Dalian, China). The target sequences were amplified in conventional nested or semi-nested PCRs. The PCR reaction was performed using TaKaRa LA Taq (Takara, Dalian, China). The overlapping PCR strategies and primers were shown in [S1 Table](#). The PCR reaction programs were shown in [S2 Table](#). The PCR products were sent to Qingke Co. (Kunming, China) for sequencing. The sequencing primers were shown in [S3 Table](#).

## Sequence analysis

The original sequences were assembled with DNA sequence analysis software Sequencher 5.0 (Gene Codes, Ann Arbor, MI). The ClustalW Multiple alignment and manual editing were performed using Bio-Edit 7.0 software. The full-length sequences dataset consisted of the eight isolates from this study and the HCV reference sequences obtained from the website of International Committee on Taxonomy of Viruses (<https://talk.ictvonline.org>), covering HCV subtypes 6a-6xf. Phylogenetic tree analyses were performed using the maximum likelihood method based on the nucleotide substitution model GTR+I+G with 1000 bootstrap replicates, using MEGA (Molecular Evolutionary Genetics Analysis, version 6.0) [14]. Pairwise nucleotide similarities/distances were calculated as uncorrected p-distances using MEGA 6.0.

## Bayesian MCMC evolutionary analyses

Based on the complete coding region sequences, Bayesian Markov chain Monte Carlo (MCMC) method was used to construct a timescaled phylogenetic tree of HCV-6. According to the previous study, the uncorrelated lognormal relaxed molecular clock model is better than the exponential and strict clock models when analyzing the HCV complete coding region sequences [15]. Under the GTR+I+ $\Gamma$ 4 nucleotide substitution model, Bayesian MCMC analyses were performed using the uncorrelated lognormal relaxed molecular clock model in combination with the Bayesian Skyline coalescent tree in the BEAST v1.8.2 package [16]. Because the dataset of HCV-6 complete coding region sequences lacked a sufficient temporal structure to estimate the evolutionary rate, an outer rate of  $1.13 \times 10^{-3} \pm 6.66 \times 10^{-6}$  substitutions/site/year was used, which was used in a previous study [17]. Each MCMC analysis was run for 50 million generations and sampled every 5,000 generations. The MCMC analysis was repeated 12 times, the resulting log-files and tree-files were combined using LogCombiner v1.8.2, respectively. The combined log-file were analyzed in Tracer v1.5, all of the ESS members were  $>200$ . The Maximum Clade Credibility (MCC) tree was obtained from the combined tree-file by TreeAnnotator v1.7.4 with a burn-in of the initial 10% of generated trees, and examined by FigTree V1.3.1, which was also used to estimate the tMRCA of various nodes on the MCC tree.

## Sequence data

The nucleotide sequences obtained in this study were submitted to GenBank under the accession numbers MK139015-MK139022.

## Results

### Demographic characteristics of the study subjects

Recently, Wan et al found a potential new HCV-6 subtype among IDUs from the China-Myanmar border area by using four partial E1E2 (KM284973-KM284976) and five partial

Table 1. The demographic characteristic of the eight isolates.

Sample ID	County	Nationality	Gender	Age	Accession Number for E1E2	Accession Number for NS5B	Accession Number for nearly full-length genome sequence
14DH34	Yingjiang	Chinese	Male	24	KT735417	KT735681	MK139015
14DH40	Longchuan	Burmese	Male	22	KT735422	KT735687	MK139016
14DH42	Longchuan	Burmese	Male	28	KT735424	KT735689	MK139017
14DH50	Yingjiang	Burmese	Male	29	KT735430	KT735697	MK139018
14DH51	Yingjiang	Burmese	Male	33	KT735431	KT735698	MK139019
14DH61	Longchuan	Burmese	Male	19	KT735438	KT735704	MK139020
14DH67	Longchuan	Burmese	Male	32	KT735442	KT735709	MK139021
14DH76	Longchuan	Burmese	Male	32	KT735444	KT735711	MK139022

<https://doi.org/10.1371/journal.pone.0217010.t001>

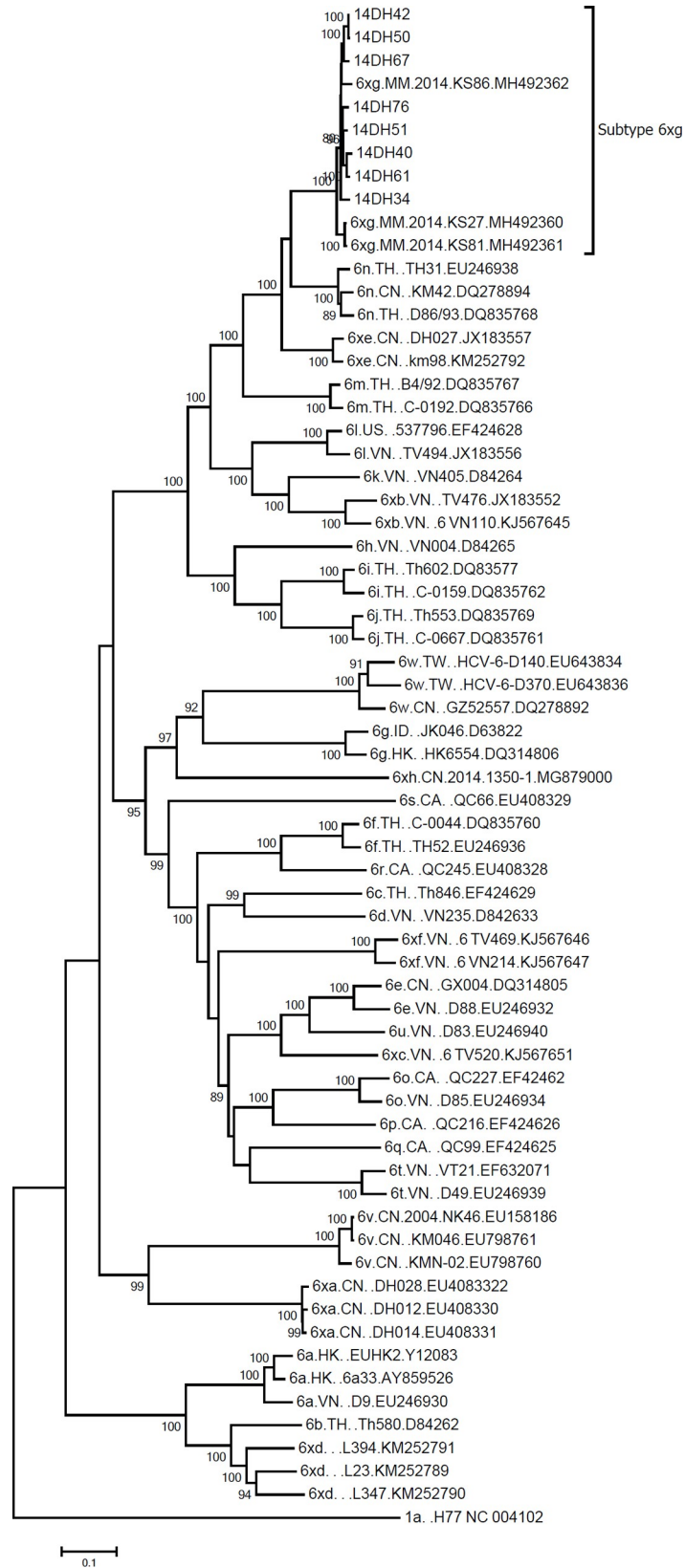
NS5B (KM285119-KM285123) sequences [18], which clustered with the eight sequences reported in one of our previous studies (Accession Numbers in Table 1) [13]. To clarify the classification of the eight strains, we further performed the complete genome sequencing and phylogenetic analysis for these strains. The eight samples were collected from eight HCV seropositive IDUs in Dehong Prefecture, bordering Myanmar. Among them, three were found in Yingjiang County, the other five were found in Longchuan County. One of them was Chinese, the other seven were Burmese. All of them were male, whose ages ranged from 19–33 years old (Table 1).

### Genome sequences and organization

Nearly full-length genome sequences of HCV were obtained from the eight HCV isolates (14DH34, 14DH40, 14DH42, 14DH50, 14DH51, 14DH61, 14DH67 and 14DH76), which were 9451–9454 nucleotides in length, crossing from 5'UTR terminus to 3'UTR and covering the complete coding region of HCV, corresponding to the nucleotides numbered 1 to 9441–9444 in the H77 genome (S4 Table). The eight genomes had the same coding region of 9051 nt, in which the 10 protein-coding regions included: core (573 nt), E1 (576 nt), E2 (1092 nt), p7 (189 nt), NS2 (651 nt), NS3 (1893 nt), NS4A (162 nt), NS4B (783 nt), NS5A (1356 nt) and NS5B (1776 nt).

### Phylogenetic analysis of full-length genome sequences

A maximum likelihood tree was constructed with the complete coding sequences of the eight isolates and 55 reference sequences of subtype 6a-6xf published on the website of International Committee on Taxonomy of Viruses (ICTV). The eight isolates formed a distinct phylogenetic clade with a strong bootstrap support, which was independent from the clades of subtype 6a-6xf and located between the clades of subtype 6n and 6xe. Then, we submitted the sequences to the ICTV Flaviviridae Study Group. After compared to the sequences in the database of the study group, the eight HCV sequences seemed to group with 6xg that was provisionally assigned in May 2018, however, whose reference sequences were not released. After the 6xg sequences (MH492360-MH492362) and 6xh sequence (MG879000) were released [19, 20], we repeated the analysis, which confirmed that the eight sequences were actually 6xg (Fig 1). Interestingly, the three 6xg sequences (MH492360-MH492362) were found in Kachin State, Myanmar [19]. Thus, the new HCV-6 subtype were reported in the three independent studies [13, 18, 19], which suggested that there were multiple distinct epidemiological unrelated isolates.



**Fig 1. The maximum likelihood phylogenetic tree of HCV complete coding region sequences.** The maximum likelihood phylogenetic tree was constructed with the complete coding region sequences of the eight isolates and the reference isolates of subtypes 6a-6xh. Values on the branches represent the percentage of 1000 bootstrap replicates. The scale bar indicates 1% nucleotide sequence divergence.

<https://doi.org/10.1371/journal.pone.0217010.g001>

### Pairwise comparison of nucleotide sequences

Based on the complete coding region, the pairwise distances of subtype 6a-6xh to these eight isolates were calculated, which ranged from 0.028 to 0.283 (Table 2). Except for subtype 6xg, subtype 6n and subtype 6xe showed less different from the eight isolates, the p-distances were 0.143–0.146 and 0.146–0.151. A genetic distance threshold to separate isolates of the same subtype from those of different subtypes is usually 13–15% [7].

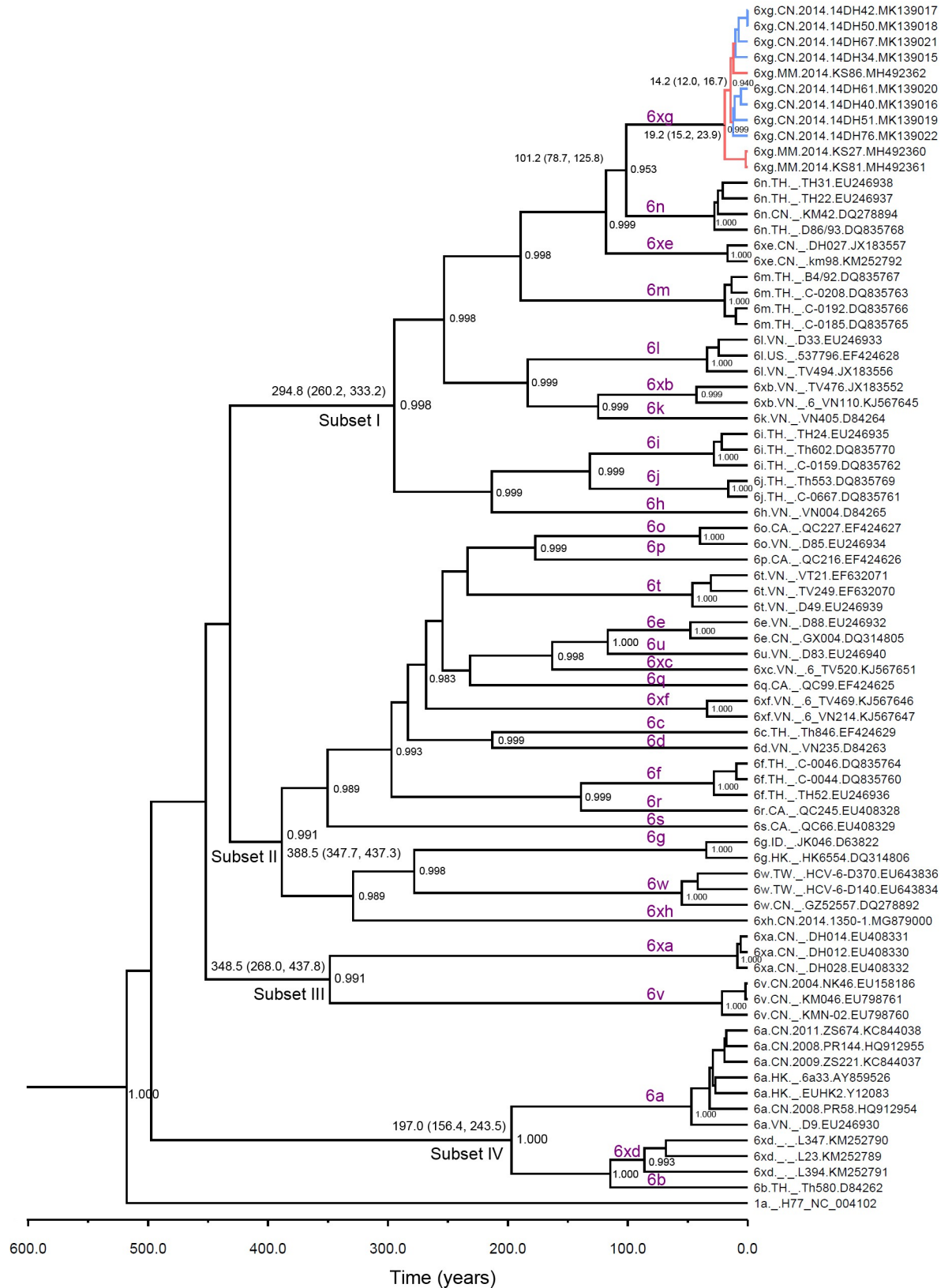
**Genetic evolutionary history of HCV-6.** With the continuous discovery of new HCV-6 subtypes, the taxonomic structure of genotype 6 continued to change. To better understand the genetic history of HCV-6, we constructed a timescaled phylogenetic tree with the complete

**Table 2. The pairwise distances of subtype 6a-6xf to the eight isolates.**

	14DH34	14DH40	14DH42	14DH50	14DH51	14DH61	14DH67	14DH76	Mean
6a_Y12083	0.279	0.283	0.282	0.283	0.281	0.283	0.281	0.279	0.281
6b_D84262	0.271	0.275	0.275	0.275	0.272	0.275	0.273	0.273	0.274
6c_EF424629	0.257	0.260	0.260	0.260	0.258	0.260	0.261	0.258	0.259
6d_D84263	0.259	0.264	0.262	0.263	0.261	0.263	0.263	0.262	0.262
6e_DQ314805	0.266	0.267	0.267	0.268	0.265	0.268	0.268	0.266	0.267
6f_DQ835760	0.263	0.265	0.264	0.265	0.263	0.263	0.265	0.263	0.264
6g_D63822	0.272	0.271	0.273	0.274	0.271	0.271	0.272	0.272	0.272
6h_D84265	0.233	0.235	0.235	0.235	0.231	0.234	0.236	0.234	0.234
6i_DQ835770	0.232	0.231	0.233	0.234	0.229	0.231	0.233	0.231	0.232
6j_DQ835769	0.232	0.232	0.235	0.235	0.231	0.232	0.234	0.232	0.233
6k_D84264	0.220	0.220	0.223	0.223	0.218	0.219	0.221	0.217	0.220
6l_EF424628	0.212	0.214	0.215	0.215	0.211	0.212	0.214	0.212	0.213
6m_DQ835767	0.185	0.188	0.187	0.188	0.187	0.188	0.187	0.187	0.187
<b>6n_DQ835768</b>	<b>0.144</b>	<b>0.146</b>	<b>0.145</b>	<b>0.145</b>	<b>0.143</b>	<b>0.145</b>	<b>0.144</b>	<b>0.143</b>	<b>0.144</b>
6o_EF424627	0.258	0.260	0.261	0.262	0.259	0.260	0.259	0.258	0.260
6p_EF424626	0.258	0.258	0.259	0.260	0.256	0.256	0.260	0.257	0.258
6q_EF424625	0.250	0.251	0.252	0.252	0.249	0.252	0.252	0.251	0.251
6r_EU408328	0.258	0.262	0.262	0.262	0.259	0.262	0.262	0.260	0.261
6s_EU408329	0.271	0.274	0.273	0.273	0.272	0.273	0.274	0.270	0.272
6t_EF632071	0.260	0.263	0.264	0.264	0.261	0.263	0.263	0.263	0.263
6u_EU246940	0.263	0.268	0.268	0.268	0.266	0.266	0.266	0.265	0.266
6v_EU158186	0.266	0.268	0.271	0.271	0.268	0.268	0.272	0.268	0.269
6w_DQ278892	0.269	0.271	0.272	0.272	0.269	0.269	0.272	0.269	0.270
6xa_EU408330	0.250	0.254	0.254	0.254	0.252	0.255	0.255	0.252	0.253
6xb_JX183552	0.222	0.222	0.225	0.225	0.220	0.220	0.224	0.221	0.222
6xc_KJ567651	0.262	0.266	0.266	0.267	0.262	0.265	0.266	0.263	0.265
6xd_KM252789	0.268	0.269	0.271	0.272	0.269	0.270	0.271	0.269	0.270
<b>6xe_JX183557</b>	<b>0.148</b>	<b>0.149</b>	<b>0.150</b>	<b>0.151</b>	<b>0.147</b>	<b>0.148</b>	<b>0.148</b>	<b>0.146</b>	<b>0.148</b>
6xf_KJ567646	0.266	0.268	0.268	0.268	0.266	0.266	0.269	0.266	0.267
6xg_MH492360	0.036	0.035	0.034	0.035	0.028	0.032	0.036	0.031	0.033
6xh_MG879000	0.265	0.266	0.267	0.267	0.266	0.266	0.267	0.266	0.266

<https://doi.org/10.1371/journal.pone.0217010.t002>





**Fig 2. Maximum clade credibility (MCC) tree representing the rooted genealogy of HCV-6 subtypes.** The branch lengths in the MCC trees reflect time and the corresponding time-scale is shown at the bottom of the trees. The posterior probabilities greater than 0.9 were shown at the respective nodes. The tMRCA for the interested nodes are shown as median (95% HPD).

<https://doi.org/10.1371/journal.pone.0217010.g002>

coding sequences of subtypes 6a-6xh. After searching the HCV database (<https://hcv.lanl.gov>) and the website of International Committee on Taxonomy of Viruses (<https://talk.ictvonline.org>), there were just 90 HCV-6 nearly full-length genome sequences, among which just nine had the annotation of collection date (S5 Table). Because of the lack of a sufficient temporal structure to estimate an evolutionary rate, we had to use an outer rate as the prior when performing Bayesian MCMC analysis, which was applied in the previous work [17, 21–25].

The maximum clade credibility (MCC) tree of genotype 6 showed that subtypes 6a-6xh were divided into four phylogenetic subsets (Fig 2). The median time of most recent common ancestor (tMRCA) of Subset I dated back to 294.8 (95% highest posterior density (HPD): 260.2, 333.2) years ago, including 6h, 6i, 6j, 6k, 6l, 6m, 6n, 6xb, 6xe and 6xg. The most of HCV-6 subtypes belonged to Subset II, whose median tMRCA dated back to 388.5 (95% HPD: 347.7, 437.3) years ago, including 6c, 6d, 6e, 6f, 6g, 6o, 6p, 6q, 6r, 6s, 6t, 6u, 6w, 6xc, 6xf and 6xh. The median tMRCA of Subset III dated back to 348.5 (95% HPD: 268.0, 437.8), just including 6v and 6xa. Subset IV was the youngest, whose median tMRCA dated back to 197.0 (95% HPD: 156.4, 243.5) years ago, including 6a, 6b and 6xd.

As shown in Fig 2, subtype 6xg clustered into Subset I, and had the most recent common ancestor with subtype 6n, which dated back to 101.2 (95% HPD: 78.7, 125.8) years ago. In the clade of subtype 6xg, the isolates from Myanmar (MH492360 and MH492361) were direct descendants of the most recent 6xg common ancestor dated back to 19.2 years ago. Branched from the isolates from Myanmar, a clade including isolates from Dehong and Myanmar descended from a common ancestor dated back to 14.2 years ago. These suggested that 6xg originated in Myanmar and transmitted into Dehong.

## Discussion

In 2005, the consensus proposal for nomenclature of HCV genotypes and subtypes was issued, which confirmed the assignment of six genotypes and 18 subtypes [8]. Because the lack of a complete coding region sequence or additional isolates, the other 58 subtypes were provisionally assigned. Based on the consensus proposal, to confirm whether an HCV isolate is a new genotype or subtype, at least one complete coding region sequence should be obtained, and the followed criteria should be met: 1) Form a distinct phylogenetic clade from previously confirmed sequences; b) At least three epidemiologically unlinked isolates are identified; c) Does not represent a recombinant between other genotypes or subtypes. After that, 38 provisional subtypes were confirmed [7]. However, the proposal has been complicated by the increasing volume of HCV sequences. In 2014, the genotypes and subtypes assignments and the nomenclature rules were updated [7], however, the criteria for assigning genotypes and subtypes remain unchanged. The changes included: a provisional subtype is not assigned without a complete genome sequence; how to assign the subtypes numbering beyond “w”; and how to name the intergenotypic recombinant forms. Meanwhile, the ICTV Flaviviridae Study Group is responsible to take a coordinating role in the assignment of newly described variants of HCV. To June 2017, the number of confirmed subtypes increased to 86.

The eight HIV-6 isolates characterized in this study were collected in 2014 in Dehong, Yunnan, China, whose partial E1E2 and NS5B sequences were firstly misclassified as subtype 6n [13]. Soon after, Wan et al found that four E1E2 and five NS5B sequences from IDUs in 2011–2012 could be a potential new HCV-6 subtype, interestingly, which clustered with the



counterparts of the eight isolates reported in our previous study [18]. Because of the lack of the complete genome sequences, the potential new HCV-6 subtype could not be assigned. Thus, we further characterized the full-length genomes of the eight HCV-6 isolates collected in 2014. The complete genome sequences of the eight isolates formed an independent clade in phylogenetic tree, and cluster with 6xg (MH492360-MH492362) that was provisionally assigned in May 2018 and released in November 13 2018. The three 6xg sequences (MH492360-MH492362) were found in Kachin State, Myanmar by Ye et al [19]. So far, the new HCV-6 subtype were found in three independent studies [13, 18, 19], which suggested that there were multiple distinct epidemiological unrelated isolates. The present study further confirmed the assignment of 6xg.

Among the eight HCV genotypes, genotype 6 is more diverse than the other seven genotypes, which suggests that HCV-6 is to represent an oldest HCV lineage [26–29]. The number of confirmed HCV-6 subtypes had increased from six in 2005 to 29 as of June 2017, accounting for one-third of all confirmed subtypes [7, 8]. However, the number of HCV-6 complete genome sequences was limited, there were just 90 HCV-6 nearly full-length genome sequences which could be found in the HCV database (<https://hcv.lanl.gov>) and the website of International Committee on Taxonomy of Viruses (<https://talk.ictvonline.org>). Among them, there were just nine with the annotation of collection date (S5 Table), which is important to construct the temporal structure when performing evolutionary analysis [15, 30]. Although the partial gene sequences were also used for evolutionary analysis, the complete coding region sequences could maximize the phylogenetic resolution [15]. Thus, the more nearly full-length genome sequences with history archives are necessary to reveal HCV evolutionary history. The lack of a sufficient temporal structure to estimate an evolutionary rate is a long-standing problem in the field of HCV research [15, 17, 21–25, 31]. The previous studies used an outer evolutionary rate as the prior to perform Bayesian MCMC analysis, which opens the door to investigate the epidemic history of HCV [15, 17, 21–25, 31]. The evolutionary rate used in this study was also utilized to analyze multiple HCV subtypes in the previous studies [17]. By this way, for the first time, we preliminarily constructed the evolutionary history with the complete coding region sequences of the HCV-6 family. A previous study showed that the approximate tMRCA of the complete coding region sequences from HCV-6 subfamily was 300–400 years ago [17], which coincided with the timescale found in our study. In future, to improve the time resolution, it is necessary to exactly estimate the evolutionary rate with the more nearly full-length genome sequences well annotated. So far, subtype 6xg nearly full-length genome sequences were annotated best among HCV-6 subtypes. We believe that ICTV Flaviviridae Study Group would get this work forward.

Southeast Asia is the ancestral and endemic region of HCV-6. The most of HCV-6 subtypes were identified in this region, including Vietnam, Thailand, Laos, Myanmar and Yunnan Province in China, where HCV-6 showed very high variability [13, 22, 27, 28, 32–36]. In this study, we found that there were four phylogenetic subsets of HCV-6 subtypes circulating in Southeast Asia. Each phylogenetic subset contained HCV-6 strains coming from the different countries, which suggested that HCV-6 had a complicated evolutionary history in this area, and at least some historical gene flow occurred. Furthermore, some newly identified HCV-6 subtypes had already circulated, such as 6xd, whose median tMRCA dated back to 86.1 years ago. Otherwise, this region contains a proportion of novel and unclassified HCV-6 variants. HCV subtype has particular epidemiological implications. As genetic markers, HCV subtypes are used to trace the transmission and evolution of HCV in specific population or areas. Thus, it is necessary to further characterize HCV-6 subtypes in this region.

According to Hepatitis C Guidance 2018 Update [37], for treatment-naïve patients with Genotype 5 and Genotype 6 infection, three regimens are recommended, namely glecaprevir/

pibrentasvir, sofosbuvir/velpatasvir and ledipasvir/sofosbuvir. To date, there is no special regimen for a certain HCV-6 subtype. However, the nearly full-length genome sequences of novel HCV-6 subtypes provide the basic for the research about the interaction between HCV and direct-acting antiviral agents (DAAs), such as naturally occurring drug resistance.

The previous studies showed that HIV-1 subtypes B and C were first introduced into IDUs in Dehong which borders Myanmar [38, 39], and further generated a number of BC recombinant forms [40]. Our previous study also showed that HCV subtype 1a, 6n and 6u among IDUs transmitted from Dehong to the other part of Yunnan, which suggested that Dehong was an entry for these subtypes [13]. In this study, the subtype 6xg was mostly identified among Burmese IDUs staying in Dehong, which suggested that this variant could come from Myanmar. The genetic evolutionary analysis further confirmed that subtype 6xg originated from Myanmar, and transmitted to Dehong through cross-border IDUs. With the development of economics, a certain amount of Burmese commuted across the China-Myanmar border, which increased the possibility of the introduction of pathogen [41, 42]. The blood-borne diseases prevention and control in border areas should be further concerned.

In this study, we characterized the nearly full-length genome sequences of subtype 6xg collected, which confirmed the circulation of subtype 6xg in China, and supported the assignment of 6xg. The genetic evolutionary analysis depicted the whole picture of the enlarged HCV-family, and improved our knowledge about the evolutionary histories of HCV-6 subtypes.

## Supporting information

### S1 Table. PCR primers.

(PDF)

### S2 Table. PCR programs.

(PDF)

### S3 Table. Sequencing primers.

(PDF)

### S4 Table. The number of nucleotides in each genomic region ORF.

(PDF)

### S5 Table. The Nearly full-length genome sequences of HCV-6 subtypes.

(PDF)

## Author Contributions

**Conceptualization:** Min Chen, Yanling Ma, Manhong Jia, Zhizhong Song.

**Data curation:** Min Chen.

**Formal analysis:** Min Chen.

**Funding acquisition:** Min Chen.

**Investigation:** Min Chen, Huichao Chen, Jie Dai, Hongbing Luo.

**Methodology:** Min Chen.

**Writing – original draft:** Min Chen.

**Writing – review & editing:** Yanling Ma, Manhong Jia, Zhizhong Song.

## References

1. Disease GBD, Injury I, Prevalence C. Global, regional, and national incidence, prevalence, and years lived with disability for 310 diseases and injuries, 1990–2015: a systematic analysis for the Global Burden of Disease Study 2015. *Lancet*. 2016; 388(10053):1545–602. [https://doi.org/10.1016/S0140-6736\(16\)31678-6](https://doi.org/10.1016/S0140-6736(16)31678-6) PMID: 27733282.
2. Lauer GM, Walker BD. Hepatitis C virus infection. *N Engl J Med*. 2001; 345(1):41–52. Epub 2001/07/07. <https://doi.org/10.1056/NEJM200107053450107> PMID: 11439948.
3. Choo QL, Kuo G, Weiner AJ, Overby LR, Bradley DW, Houghton M. Isolation of a cDNA clone derived from a blood-borne non-A, non-B viral hepatitis genome. *Science*. 1989; 244(4902):359–62. PMID: 2523562.
4. Webster DP, Klenerman P, Dusheiko GM. Hepatitis C. *Lancet*. 2015; 385(9973):1124–35. [https://doi.org/10.1016/S0140-6736\(14\)62401-6](https://doi.org/10.1016/S0140-6736(14)62401-6) PMID: 25687730.
5. Messina JP, Humphreys I, Flaxman A, Brown A, Cooke GS, Pybus OG, et al. Global distribution and prevalence of hepatitis C virus genotypes. *Hepatology*. 2015; 61(1):77–87. <https://doi.org/10.1002/hep.27259> PMID: 25069599.
6. Gower E, Estes C, Blach S, Razavi-Shearer K, Razavi H. Global epidemiology and genotype distribution of the hepatitis C virus infection. *J Hepatol*. 2014; 61(1 Suppl):S45–57. Epub 2014/08/03. <https://doi.org/10.1016/j.jhep.2014.07.027> PMID: 25086286.
7. Smith DB, Bukh J, Kuiken C, Muerhoff AS, Rice CM, Stapleton JT, et al. Expanded classification of hepatitis C virus into 7 genotypes and 67 subtypes: updated criteria and genotype assignment web resource. *Hepatology*. 2014; 59(1):318–27. Epub 2013/10/12. <https://doi.org/10.1002/hep.26744> PMID: 24115039.
8. Simmonds P, Bukh J, Combet C, Deleage G, Enomoto N, Feinstone S, et al. Consensus proposals for a unified system of nomenclature of hepatitis C virus genotypes. *Hepatology*. 2005; 42(4):962–73. Epub 2005/09/09. <https://doi.org/10.1002/hep.20819> PMID: 16149085.
9. Pybus OG, Barnes E, Taggart R, Lemey P, Markov PV, Rasachak B, et al. Genetic history of hepatitis C virus in East Asia. *J Virol*. 2009; 83(2):1071–82. Epub 2008/10/31. <https://doi.org/10.1128/JVI.01501-08> PMID: 18971279.
10. Lu L, Jia M, Ma Y, Yang L, Chen Z, Ho DD, et al. The changing face of HIV in China. *Nature*. 2008; 455(7213):609–11. <https://doi.org/10.1038/455609a> PMID: 18833270.
11. Wu Z, Sullivan SG, Wang Y, Rotheram-Borus MJ, Detels R. Evolution of China's response to HIV/AIDS. *Lancet*. 2007; 369(9562):679–90. Epub 2007/02/27. [https://doi.org/10.1016/S0140-6736\(07\)60315-8](https://doi.org/10.1016/S0140-6736(07)60315-8) PMID: 17321313.
12. Lu L, Murphy D, Li C, Liu S, Xia X, Pham PH, et al. Complete genomes of three subtype 6t isolates and analysis of many novel hepatitis C virus variants within genotype 6. *J Gen Virol*. 2008; 89(Pt 2):444–52. <https://doi.org/10.1099/vir.0.83460-0> PMID: 18198375.
13. Chen M, Ma Y, Chen H, Luo H, Dai J, Song L, et al. Multiple Introduction and Naturally Occurring Drug Resistance of HCV among HIV-Infected Intravenous Drug Users in Yunnan: An Origin of China's HIV/HCV Epidemics. *PLoS One*. 2015; 10(11):e0142543. Epub 2015/11/13. <https://doi.org/10.1371/journal.pone.0142543> PMID: 26562015.
14. Tamura K, Stecher G, Peterson D, Filipinski A, Kumar S. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Mol Biol Evol*. 2013; 30(12):2725–9. <https://doi.org/10.1093/molbev/mst197> PMID: 24132122.
15. Yuan M, Lu T, Li C, Lu L. The evolutionary rates of HCV estimated with subtype 1a and 1b sequences over the ORF length and in different genomic regions. *PLoS One*. 2013; 8(6):e64698. Epub 2013/06/14. <https://doi.org/10.1371/journal.pone.0064698> PMID: 23762247.
16. Drummond AJ, Suchard MA, Xie D, Rambaut A. Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Molecular biology and evolution*. 2012; 29(8):1969–73. <https://doi.org/10.1093/molbev/mss075> PMID: 22367748.
17. Lu L, Li C, Xu Y, Murphy DG. Full-length genomes of 16 hepatitis C virus genotype 1 isolates representing subtypes 1c, 1d, 1e, 1g, 1h, 1i, 1j and 1k, and two new subtypes 1m and 1n, and four unclassified variants reveal ancestral relationships among subtypes. *J Gen Virol*. 2014; 95(Pt 7):1479–87. <https://doi.org/10.1099/vir.0.064980-0> PMID: 24718832.
18. Wan Z, Chen Q, Chen X, Duo L, Li P, Zheng YT, et al. HCV Diversity among Chinese and Burmese IDUs in Dehong, Yunnan, China. *PLoS One*. 2016; 11(9):e0163062. <https://doi.org/10.1371/journal.pone.0163062> PMID: 27657722.
19. Ye M, Chen X, Wang Y, Duo L, Zhang C, Zheng Y-T. Identification of a new HCV subtype 6xg among injection drug users in Kachin, Myanmar. *Frontiers in Microbiology*. 2019; 10:814. <https://doi.org/10.3389/fmicb.2019.00814> PMID: 31057519

20. Wu T, Xing Z, Yuan M, Ge J, Yuan G, Liang K, et al. Analysis of HCV Isolates Among the Li Ethnic in Hainan Island of South China Reveals Their HCV-6 Unique Evolution and a New Subtype. *Cell Physiol Biochem*. 2018; 50(5):1832–9. Epub 2018/11/06. <https://doi.org/10.1159/000494863> PMID: 30396187.
21. Lu L, Wang M, Xia W, Tian L, Xu R, Li C, et al. Migration patterns of hepatitis C virus in China characterized for five major subtypes based on samples from 411 volunteer blood donors from 17 provinces and municipalities. *J Virol*. 2014; 88(13):7120–9. <https://doi.org/10.1128/JVI.00414-14> PMID: 24719413.
22. Li C, Yuan M, Lu L, Lu T, Xia W, Pham VH, et al. The genetic diversity and evolutionary history of hepatitis C virus in Vietnam. *Virology*. 2014;468–470C:197–206. <https://doi.org/10.1016/j.virol.2014.07.026> PMID: 25193655.
23. An Y, Wu T, Wang M, Lu L, Li C, Zhou Y, et al. Conservation in China of a novel group of HCV variants dating to six centuries ago. *Virology*. 2014;464–465:21–5. Epub 2014/07/22. <https://doi.org/10.1016/j.virol.2014.06.011> PMID: 25043585.
24. Njouom R, Frost E, Deslandes S, Mamadou-Yaya F, Labbe AC, Pouillot R, et al. Predominance of hepatitis C virus genotype 4 infection and rapid transmission between 1935 and 1965 in the Central African Republic. *J Gen Virol*. 2009; 90(Pt 10):2452–6. Epub 2009/06/19. <https://doi.org/10.1099/vir.0.011981-0> PMID: 19535500.
25. Pouillot R, Lachenal G, Pybus OG, Rousset D, Njouom R. Variable epidemic histories of hepatitis C virus genotype 2 infection in West Africa and Cameroon. *Infect Genet Evol*. 2008; 8(5):676–81. Epub 2008/07/05. <https://doi.org/10.1016/j.meegid.2008.06.001> PMID: 18599380.
26. Salemi M, Vandamme AM. Hepatitis C virus evolutionary patterns studied through analysis of full-genome sequences. *J Mol Evol*. 2002; 54(1):62–70. Epub 2001/12/06. <https://doi.org/10.1007/s00239-001-0018-9> PMID: 11734899.
27. Wang H, Yuan Z, Barnes E, Yuan M, Li C, Fu Y, et al. Eight novel hepatitis C virus genomes reveal the changing taxonomic structure of genotype 6. *J Gen Virol*. 2013; 94(Pt 1):76–80. Epub 2012/09/28. <https://doi.org/10.1099/vir.0.047506-0> PMID: 23015745.
28. Li C, Barnes E, Newton PN, Fu Y, Vongsouvath M, Klenerman P, et al. An expanded taxonomy of hepatitis C virus genotype 6: Characterization of 22 new full-length viral genomes. *Virology*. 2015; 476:355–63. <https://doi.org/10.1016/j.virol.2014.12.025> PMID: 25589238.
29. Li C, Pham VH, Abe K, Lu L. Nine additional complete genome sequences of HCV genotype 6 from Vietnam including new subtypes 6xb and 6xc. *Virology*. 2014;468–470:172–7. Epub 2014/09/01. <https://doi.org/10.1016/j.virol.2014.08.006> PMID: 25173194.
30. Pybus OG, Charleston MA, Gupta S, Rambaut A, Holmes EC, Harvey PH. The epidemic behavior of the hepatitis C virus. *Science*. 2001; 292(5525):2323–5. Epub 2001/06/26. <https://doi.org/10.1126/science.1058321> PMID: 11423661.
31. Cheng H, Zhang J, Capizzi J, Young NL, Mastro TD. HIV-1 subtype E in Yunnan, China. *Lancet*. 1994; 344(8927):953–4. PMID: 7934363.
32. Xia X, Zhao W, Tee KK, Feng Y, Takebe Y, Li Q, et al. Complete genome sequencing and phylogenetic analysis of HCV isolates from China reveals a new subtype, designated 6u. *J Med Virol*. 2008; 80(10):1740–6. <https://doi.org/10.1002/jmv.21287> PMID: 18712831.
33. Shinji T, Kyaw YY, Gokan K, Tanaka Y, Ochi K, Kusano N, et al. Analysis of HCV genotypes from blood donors shows three new HCV type 6 subgroups exist in Myanmar. *Acta medica Okayama*. 2004; 58(3):135–42. <https://doi.org/10.18926/AMO/32110> PMID: 15471435.
34. Shinji T, Lwin AA, Gokan K, Obika M, Ryuko H, Khin M, et al. Three type 6 hepatitis C virus subgroups among blood donors in the Yangon area of Myanmar are identified as subtypes 6m and 6n, and a novel subtype by sequence analysis of the core region. *Acta medica Okayama*. 2006; 60(6):345–9. <https://doi.org/10.18926/AMO/30719> PMID: 17189978.
35. Hubschen JM, Jutavijittum P, Thammavong T, Samouny B, Yousukh A, Toriyama K, et al. High genetic diversity including potential new subtypes of hepatitis C virus genotype 6 in Lao People's Democratic Republic. *Clin Microbiol Infect*. 2011; 17(12):E30–4. <https://doi.org/10.1111/j.1469-0691.2011.03665.x> PMID: 21958219.
36. Pham VH, Nguyen HD, Ho PT, Banh DV, Pham HL, Pham PH, et al. Very high prevalence of hepatitis C virus genotype 6 variants in southern Vietnam: large-scale survey based on sequence determination. *Japanese journal of infectious diseases*. 2011; 64(6):537–9. PMID: 22116339.
37. Panel A-IHG. Hepatitis C Guidance 2018 Update: AASLD-IDS Recommendations for Testing, Managing, and Treating Hepatitis C Virus Infection. *Clin Infect Dis*. 2018; 67(10):1477–92. Epub 2018/09/15. <https://doi.org/10.1093/cid/ciy585> PMID: 30215672.
38. Graf M, Shao Y, Zhao Q, Seidl T, Kostler J, Wolf H, et al. Cloning and characterization of a virtually full-length HIV type 1 genome from a subtype B'-Thai strain representing the most prevalent B-clade isolate in China. *AIDS Res Hum Retroviruses*. 1998; 14(3):285–8. <https://doi.org/10.1089/aid.1998.14.285> PMID: 9491920.

39. Luo CC, Tian C, Hu DJ, Kai M, Dondero T, Zheng X. HIV-1 subtype C in China. *Lancet*. 1995; 345 (8956):1051–2. PMID: [7723522](#).
40. Feng Y, Takebe Y, Wei H, He X, Hsi JH, Li Z, et al. Geographic origin and evolutionary history of China's two predominant HIV-1 circulating recombinant forms, CRF07\_BC and CRF08\_BC. *Sci Rep*. 2016; 6:19279. <https://doi.org/10.1038/srep19279> PMID: [26763952](#).
41. Chen X, Duo L, Ye M, Zhang C, Zheng YT. Non-Chinese Immigrants: Challenge Faced by Yunnan of China to Achieve the 90-90-90 Goals. *Virology*. 2018; 33(3):291–3. Epub 2018/06/23. <https://doi.org/10.1007/s12250-018-0038-x> PMID: [29931513](#).
42. Chen X, Zhou YH, Ye M, Wang Y, Duo L, Pang W, et al. Burmese injecting drug users in Yunnan play a pivotal role in the cross-border transmission of HIV-1 in the China-Myanmar border region. *Virulence*. 2018; 9(1):1195–204. Epub 2018/07/13. <https://doi.org/10.1080/21505594.2018.1496777> PMID: [30001176](#).