Original Paper

Genotype Distribution and Molecular Epidemiology of Hepatitis C Virus in Guangzhou, China: Predominance of Genotype 1b and Increasing Incidence of Genotype 6a

Guosheng Yuan^a Junwei Liu^a Chengguang Hu^a Huaping Huang^a
Minghua Qi^{a,b} Tao Wu^{a,c} Weifang Liang^a Yi-Ping Li^d Yong-Yuan Zhang^e
Yuanping Zhou^a

^aDepartment of Infectious Diseases and Hepatology Unit, Nanfang Hospital, Southern Medical University, Guangzhou, ^bDepartment of Infectious Diseases, Peking University Shenzhen Hospital, Shenzhen, ^cDepartment of Infectious Disease, Hainan General Hospital, Haikou, ^dInstitute of Human Virology and Key Laboratory of Tropical Disease Control of Ministry of Education, Zhongshan School of Medicine, Sun Yat-sen University, Guangzhou, China; ^eHBVtech, Germantown, Maryland, MD, USA

Key Words

Molecular epidemiology • HCV • Evolution • Phylogeographic analysis • China • Guangdong province

Abstract

Background/Aims: Distribution of Hepatitis C virus (HCV) genotypes vary geographically and may associate with the mode of transmission. Little is known about the molecular epidemiology of HCV infection in Guangzhou, China. **Methods:** A cross-sectional survey included 561 subjects with chronic HCV infection registered at Nanfang Hospital, Southern Medical University, was performed. All residents were invited for a questionnaire interview to collect information about their personal status and commercial blood donation history. **Results:** A total of 463 chronic hepatitis C (CHC) patients were finally enrolled. Among the 463 samples, 426 were characterized by partial core-E1 sequences and classified into 7 subtypes: 1b (n=263, 61.7%), 6a (n=86, 20.2%), 2a (n=26, 6.1%), 3b (n=26, 6.1%), 3a (n=22, 5.2%), 6u (n=2, 0.5%), and 4a (n=1, 0.2%). Analysis of genotype-associated risk factors revealed that blood donation and transfusion were strongly associated with subtypes 1b and 2a, while genotype 3b and 6a were more frequent in intravenous drug users. **Conclusions:** Phylogeographic analyses demonstrated that the distribution of HCV genotypes in Guangzhou is complex. Interestingly, 6a has become a local endemic in Guangzhou and may be the second source region to disseminate 6a to other provinces.

© 2017 The Author(s)
Published by S. Karger AG, Basel

Yuanping Zhou

Department of Infectious Diseases and Hepatology Unit, Nanfang Hospital,
Southern Medical University, Guangzhou (China)
E-Mail yuanpingzhou@163.com

KARGER

Introduction

As a worldwide health problem, an estimated 130-150 million people are chronically infected with Hepatitis C Virus (HCV), or 2.5% of the global population [1-3]. Taxonomically, HCV is classified into seven confirmed genotypes, and each genotype, except for genotypes 5 and 7, is further divided into a number of subtypes [4]. Different genotypes have shown distinct geographic distribution patterns. In general, genotypes 1, 2, and 3 are prevalent worldwide, while genotypes 4 and 5 are primarily restricted to Africa and genotype 6 is endemic to Southeast Asia [1, 3, 5, 6]. However, such patterns are constantly evolving as a result of speedy transmission via global travel.

Although six genotypes (genotypes 1 to 6) and a number of subtypes have been detected in China, over 95% of these isolates belong to five major subtypes: 1b, 2a, 3a, 3b, and 6a [5, 7-11]. Among them, subtype 1b is predominant nationwide, accounting for approximately 75% of all HCV infections, followed by 2a [5, 10, 11]. However, little is known about the distribution pattern of HCV genotypes in Guangzhou, the capital of Guangdong province, China (Fig. 1).

Guangzhou is not only the capital of Guangdong, but also the central city with strong attraction and influence in South China. Known as one of the starts of the ancient Marine Silk Road of China and one of the sources of modern China revolution, it has become the frontline of the reform and opening-up of China since 1978. Recently, as its brilliant achievements in developing economy and improving urban aspects, Guangzhou has become a “World

Fig. 1. Map of HCV 1b and 6a migration in China. Being a “World Production Center”, Guangzhou as the core of the Pearl River Delta metropolitan area could be the origin of 1b subtype in China (shown in red circle and purple arrows). With origin from Vietnam, 6a was spread to the neighboring Guangxi and Yunnan Provinces for circulation. From Guangxi, 6a was further spread to the Guangzhou, and became a local epidemic. As the prevalence is increasing, Guangzhou has become the second source region to transmit the 6a to other regions (shown in pink and purple arrows). (The software Google Earth (Map data: Google, DigitalGlobe) was used to create the map.)

Production Center". However, this has also brought about many side effects, such as the increasing drug use, drug trafficking, prostitution, unsafe medical practices, and millions of migrant laborers living in suboptimal hygiene conditions. All of these have contributed to create an environment to inbreed HCV prevalence. Here we utilized samples collected from Chronic Hepatitis C (CHC) patients in Guangzhou to determine the status of HCV infection at this region, and related risk factors, as well as its relationship with those in other regions of China.

Materials and Methods

Subjects and specimens

Participants were recruited from July 2007 to December 2015 in Nanfang Hospital, Southern Medical University in Guangzhou, Guangdong Province, China. A total of 463 patients were finally enrolled in the study and all of them had anti-HCV positivity and detectable serum HCV RNA for at least 6 months. Blood samples were centrifuged and the supernatants were stored at -80°C for HCV genotyping. Laboratory results were collected from clinical records and the additional information about their personal status and commercial blood donation history was obtained using a standardized questionnaire. The serum HCV DNA level was measured with the Cobas AmpliPrep/Cobas TaqMan HCV test, version 2.0 (CAP/CTM HCV v2.0; lower limit of detection, 15 IU/mL). Possible HCV infection sources were identified for each patient including transfusion, intravenous drug use (IDU) and other procedures that share the equipment such as hemodialysis, dental treatment, non-sterile tattooing or piercing, unsafe acupuncture, etc.

The study was approved by the ethical Committee of Nanfang Hospital, Southern Medical University. The experiments were carried out in accordance with the approved guidelines and the "informed" consent was obtained from all subjects.

HCV RNA extraction, RT-PCR amplification, and sequencing

Serum HCV RNA was extracted from 140 µl of plasma by using QIAamp Viral RNA mini Kit (Qiagen, Germany) according to the recommended protocol. cDNA was synthesized from 20 µl extracted RNA with Superscript III-First-Strand Synthesis System (Invitrogen, Life Technologies, USA). The primers of the partial Core-E1 region (reference strain H77 positions: 729–1322nt) for nested-PCR were described elsewhere: outer forward (E1) 5'-TTGGGTAAGGTCATCGATACCC-3', outer reverse (E2) 5'-TGATGTGCCAACTGCCGTTGGT-3'; inner forward (E3) 5'-TTCGCCGACCTCATGGGGTACAT-3', inner reverse (E4) 5'-GGACCAGTTCATCATCATATCCCA-3'. First round PCR using primers E1 and E2 was conducted with the following conditions: 94°C for 2 min, 35 cycles of 94°C for 30sec, 58°C for 1min and 72°C for 40sec, then a final cycle at 72°C for 7min. Second round PCR using primers E3 and E4 was conducted with the same condition except that the annealing temperature was 56°C for 35 sec. The amplicons were identified on 1.0% agarose gel, and directly subjected to sequencing by Shanghai Invitrogen Biotechnology Co., Ltd. To avoid potential contamination, experimental procedures were strictly performed by adding positive and negative controls and processed in parallel, including extraction of RNA, preparation of reagents, RT-PCR, nested PCR, and gel electrophoresis. Sequencing was done in both directions with primers E3 and E4 using ABI Prism Big Dye 3.0 terminators on an ABI Prism 3500 genetic analyzer (PE Applied Biosystems, Foster City, CA, USA).

HCV genotyping and phylogenetic analyses

The nucleotide sequences obtained were aligned with HCV strains of standard genotypes and edited by the ClustalW method of MEGA software (Version 6.0). Prior to phylogenetic tree construction, the best-fitting substitution model was tested using the jModeltest program (version 2.1.7) on the basis of the Akaike Information Criterion, which demonstrated that GTR+I+ Γ (GTR + invariant sites + gamma rate heterogeneity) was the best model for all of the sequence datasets. Under this model, the maximum likelihood (ML) trees were heuristically searched using the subtree pruning and regrafting (SPR) algorithm and the nearest-neighbor interchange (NNI) perturbation algorithm implemented in PhyML software, with which bootstrap analyses were performed in 500 replicates. After NEXUS tree files were generated, the ML tree topology was displayed using the FigTree program.

Phylogeographic tree analysis

Phylogeographic trees were reconstructed using the Bayesian phylogeographic inference framework implemented in the BEAST software (version 2.2.1), where a Bayesian discrete phylogeographic approach and a Bayesian Stochastic Search Variable Selection (BSSVS) procedure were used to estimate the ancestral locations of the virus and infer the most significant epidemiological links, respectively. Before constructing the trees, site models, demographic models, clock models, and evolutionary rates were estimated using the BEAST package, which involved all sequences obtained in this study as well as reference sequences. Briefly, the combination of the GTR+I+Γ4 substitution model, the Bayesian skyline coalescent model, and the uncorrelated exponential clock mode was selected because this combination always outperformed other combinations [6, 9]. With the aforementioned datasets and the BEAST settings described previously, evolutionary rates (substitution per site per year) were estimated and used as priors in the present study: $1.51 \times 10^{-3} \pm 2.66 \times 10^{-5}$, $2.78 \times 10^{-3} \pm 2.30 \times 10^{-6}$, $2.00 \times 10^{-3} \pm 3.00 \times 10^{-5}$, $5.20 \times 10^{-3} \pm 6.85 \times 10^{-5}$, and $2.74 \times 10^{-3} \pm 1.51 \times 10^{-5}$ for the subtypes 1b, 2a, 3a, 3b, 6a Core-E1 sequences, respectively [9, 12, 13]. After generating the XML files and importing them into BEAST, the Markov Chain Monte Carlo (MCMC) chain was run for 200 million states and sampled every 10,000 states, ensuring that sufficient sampling has been achieved, indicated by the estimated effective sampling sizes (ESSs) greater than 200. After that, the program TreeAnnotator was used to generate the maximum clade credibility (MCC) tree as well as to summarize the posterior density of trees to calculate the posterior probabilities for the ancestral geographic states. Moreover, Tracer was used to explore the output of BEAST and the FigTree program was applied to display the resultant posterior trees.

In addition, to test for the presence of phylogeographic structures, the Befi-BaTS program (http://lonelyjoeparker.com/wp/?page_id=274#Befi-BaTS) was performed to estimate two statistics: AI (Association Index) and PS (Parsimony Score statistic). AI is the sum across all the internal nodes of a phylogeny, which explicitly takes into account the shape of the phylogeny by measuring the imbalance of the internal nodes. The estimation of PS, however, uses a parsimony approach to reconstruct the character states at ancestral nodes and calculate the number of state changes in the phylogeny. This was done through a parametric bootstrap process, which randomizes the association a large number of times, calculates statistics from each randomization, and provides a null distribution of the statistics with 0.95 credible intervals. Against this level the observed statistics are compared. Here, the null hypothesis of panmixis assumes that there is no correlation between phylogeny and taxa location [14], and the randomization is performed against a series of tree-shaped statistics. We performed randomizations across a posterior distribution of trees generated from the MCMC process under a given coalescent model. During the bootstrapping process, the phylogenetic uncertainties were correctly incorporated and phylogeographic structure tested [15].

In this study, we have included 150 GT1b, 78 GT2a, 41 GT3b, 24 GT3a and 69 GT6a reference sequences from a single study, which covered HCV isolates collected from many different regions in Mainland China [16]. In addition, to determine how the results could be affected by subsampling, we have randomly sub-sampled the Guangzhou sequences that more or less matched the same size in different regions (unpublished data) in our analysis.

Statistical analysis

The SPSS 17 software (SPSS Inc., Chicago, IL, USA) package was used for testing differences between groups by the Chi-square test. Analysis of Variance (ANOVA) was performed to test the quantitative data (HCV RNA level). A *p*-value of <0.05 was considered statistically significant.

Nucleotide sequence accession numbers

The nucleotide sequences reported in this study were deposited in GenBank with the accession numbers: KU364624 - KU365049.

Results

Demographic and clinical characteristics

Of 463 samples, 426 (92.0%) were successfully amplified and sequenced. The ages of the patients in the study ranged from 6 to 81 years with a mean age ($\pm$ SD) of 42.6 ± 12.9 years.

Table 1. Demographics and risk factors of patients associated with HCV subtypes (1b, 2a, 3a, 3b and 6a),
*To avoid statistical bias, the HCV subtype 4a (n=1) and 6u (n=2) were excluded from the analysis because of their limited samples numbers. Blood, patients with a history of transfusion blood or blood products; IDU, patients with a history of Intravenous drug use; Sexual contact, patients with a history of promiscuity and unsafe sex; MTCT: mother-to-child transmission; Others: tattooing, piercing, acupuncture, hemodialysis, etc.; Unknown, patients where the source of infection was unclear and unknown

Factor	HCV subtype*					χ^2 / F	P value
	1b(n=263)	2a(n=26)	3a(n=22)	3b(n=26)	6a(n=86)		
Gender							
Male	152(57.8%)	10(38.5%)	18(81.8%)	19(73.1%)	61(70.9%)	16.307	0.003
Age(years)	45	47	38	40	36	5.984	0.000
Age(cases)						78.188	0.000
≤20	14(5.3%)	1(3.8%)	0	2(7.7%)	0		
21-30	36(13.7%)	2(7.7%)	3(13.6%)	3(11.5%)	9(10.5%)		
31-40	52(19.8%)	4(15.4%)	10(45.5%)	11(42.3%)	36(41.9%)		
41-50	67(25.5%)	9(34.6%)	9(40.9%)	6(23.1%)	36(41.9%)		
>50	94(35.7%)	10(38.5%)	0	4(15.4%)	5(5.8%)		
HCV RNA (lg IU/mL, mean ± SD)	5.87±1.18	5.72±0.92	5.74±1.01	5.76±0.99	5.74±1.01	2.345	0.054
Risk behavior							
Blood	198(75.3%)	15(57.7%)	8(36.4%)	2(7.7%)	11(12.8%)	177.210	0.000
IDU	23(8.7%)	3(11.5%)	3(13.6%)	17(65.7%)	34(39.5%)		
Sexual contact	11(4.2%)	1(3.8%)	2(9.1%)	2(7.7%)	1(1.2%)		
MTCT	1(0.4%)	0	0	1(3.8%)	0		
Others	13(4.9%)	3(11.5%)	4(18.2%)	3(11.5%)	24(27.9%)		
unknown	17(6.5%)	4(15.4%)	5(22.7%)	1(3.8%)	16(18.6%)		

Fig. 2. Circular form of a phylogenetic tree based on the partial C/E1 sequences amplified from the 426 CHC patients. Different subtypes are shown in different colors, as indicated on the left of the tree. A vertical ruler with a length of 0.1 nucleotides per site is shown in the middle of the tree as a guide to measure the genetic distances. The pie chart inside the tree indicates the percentages of the different HCV subtypes (indicated in the same color coding used for the circular tree) into which the 426 CHC patients were classified.

Of these, 263 patients (61.7%) were male and 163 patients (38.3%) were female (Table 1).

No differences were found in serum HCV RNA levels among the HCV genotypes (Table 1); mean HCV RNA levels (IU/mL) in log10 were: 5.87 ± 1.18 for genotype 1b, 5.72 ± 0.92 for genotype 2a, 5.74 ± 1.01 for genotype 3a, 5.76 ± 0.99 for genotype 3b, and 5.74 ± 1.01 for genotype 6.

Fig. 3. Phylogeographic tree estimated with the partial Core-E1 sequences of subtype 1b isolates. A total of 413 subtype 1b sequences were used to generate the phylogeographic tree: 263 obtained in this study, and 150 from blood donors from the 17 provinces and municipalities in China. Branches are colored according to their geographic origins, indicated in the upper left. Posterior probabilities of >0.70 are shown at the respective nodes. Below the tree is a time scale from 1955 to 2015, which indicates the time of HCV origin and evolution.

HCV genotype distribution

After aligning and editing with HCV strains of standard genotypes, a 376 bp of the partial Core-E1 region (H77 positions: 892–1267nt) was utilized for HCV genotyping and phylogenetic analyses. The all sequences were classified into seven HCV subtypes: 1b, 2a, 3a, 3b, 4a, 6a, 6u (Fig. 2). HCV subtypes 1b was the most prevalent subtype ($n = 263$, 61.7%), followed by subtype 6a ($n = 86$, 20.2%), 2a ($n = 26$, 6.1%), 3b ($n = 26$, 6.1%) and 3a ($n = 22$, 5.2%), while subtype 4a and 6u accounted for HCV infection in only 1 (0.2%) and 2 (0.5%) individuals, respectively.

Furthermore, the main HCV genotype distribution differed significantly by gender. Subtype 1b (men/women: 57.8% vs. 42.2%), 3a (81.8% vs. 18.2%), 3b (73.1% vs. 26.9%) and 6a (70.9% vs. 29.1%) were more frequent in men than in women. In contrast, subtype 2a (men/women: 38.5% vs. 61.5%) were more frequent in women than in men ($\chi^2 = 16.307$, $P = 0.003$). Besides, patients infected with genotype 1b and 2a presented the increasing tendency with age comparing with other genotypes ($\chi^2 = 78.188$, $P < 0.001$).

More importantly, the distributions of HCV genotypes showed significant differences depending on the route of transmission ($P < 0.001$). Genotype 1b was observed more frequently (75.3%) in patients infected through blood transfusion than other routes, while genotype 3b (65.7%) and 6a (39.5%) were more frequent in those infected through IDU (Table 1).

Phylogeographic analysis

Briefly, five trees were generated, each representing one of the five common subtypes in Guangzhou, i.e., 1b, 2a, 3a, 3b, and 6a.

Fig. 3 presents a phylogeographic tree reconstructed on the basis of the 413 Core-E1 sequences of subtype 1b. Six groups are indicated as group A, B, C, D, E and F, showing a significant

posterior probability of 1.00. Group A sequences mainly consists of a single region, the northwest. The strains in Group A may have originated from a Guangzhou ancestor dated around 1982 (95% CI: 1979–1985). Between Group A and B, there appears to be two additional groups, one of which contains sequences mostly from Guangzhou and the north-northeast, but without significant posterior probabilities. Sequences in Group B, C and F were monolithically from Guangzhou, indicating that subtype 1b was locally endemic in this region. It seems that these isolates may have descended from Guangzhou ancestors dated around 1978 (95% CI: 1975–1983), 1976 (95% CI: 1962–1982) and 1982 (95% CI: 1975–1990), respectively. In contrast to the above groups, Group D and E contained sequences with a mixture of geographic origins, showing a wide range of geographic distribution and a simultaneous dissemination nationwide. Both of them may also have descended from Guangzhou ancestors, dated around 1982 (95% CI: 1980–1987) and 1972 (95% CI: 1968–1980), respectively.

Fig. 4A presents a phylogeographic tree for 104 subtype 2a sequences. Overall, these 2a sequences can be classified into two big groups, showing significant posterior probabilities of 0.91 and 1.00, respectively. The majority of sequences in Group A may originate from Guangzhou, while Group B displays branches with a mixture of geographic origins but mainly from the northwest. Upward from the base, there appears to be a trend of 2a migration first to the northwest around 1966 (95% CI: 1957–1974), then gradually from northwest to many other regions.

Fig. 4B displays a phylogeographic tree for 46 subtype 3a sequences. Compared to the two trees described above for subtypes 1b and 2a, the tree for the subtype 3a sequences appears to be clean and structured in an orderly manner, showing three clearly separated geographic groups, in addition to a single branch. Each of three groups contains sequences originating almost exclusively from a single geographic region, the southwest, northwest, or Guangzhou, and each reaches a full posterior probability of 1.00.

Fig. 4C displays a phylogeographic tree for 67 subtype 3b sequences. The tree could roughly be divided into two subsets. The smaller subset is located at the tree base and contains only two branches but shows a full posterior probability of 1.00. With a full posterior of 1.00, the larger subset appears to show a migration trend from the southwest to Guangzhou and sporadically to the northwest, with a few migrations to the north-northeast and the southeast. As a whole, 3b in China may have originated from Yunnan in the southwest, and the 3b isolates in other regions may be viewed as descendants.

The 155 subtype 6a sequences are all maintained and supported with posterior probabilities of >0.80 in Fig. 5. The base consisting of Vietnamese isolates, contains the most direct descendants of the earliest 6a common ancestor dated around 1957 (95% CI: 1939–1984), which is thought to be the origin of all 6a strains in China. Upward from the base, there appears to be a route of 6a migration: it initially reaches Guangzhou, and then spreads to many other regions.

Migration test

According to sampling origins, the GT1b, 2a, 3a and 3b sequences were divided into 6 states (North & Northeast, Northwest, Central South, Southeast, Southwest, and Guangzhou), while GT6a was divided into 6 different kind of states (Guangzhou, Guangxi, Hubei, Yunnan, Vietnam and Other provinces). Correlating the above 6 states with phylogenetic positions showed the AI and PS statistics both strongly rejecting the null hypothesis of panmixis. Therefore, association between lineages and geographic origins was supported.

Discussion

In the present study, a total of 463 HCV RNA positive samples was collected from HCV infected patients who visited Nanfang Hospital, Southern Medical University in Guangzhou, Guangdong Province, China. Of these samples, 426 (92.0%) were successfully amplified and

Fig. 4. Phylogeographic tree estimated with the partial Core-E1 sequences of subtype 2a, 3a and 3b isolates. Fig. 4A A total of 104 subtype 2a sequences were used to generate the phylogeographic tree: 26 obtained in this study, and 78 from blood donors from the 17 provinces and municipalities in China. Below the tree is a time scale from 1950 to 2015; Fig. 4B A total of 46 subtype 3a sequences were used to generate the phylogeographic tree: 22 obtained in this study, and 24 from blood donors from the 17 provinces and municipalities in China. Below the tree is a time scale from 1975 to 2015; Fig. 4C A total of 67 subtype 3b sequences were used to generate the phylogeographic tree: 26 obtained in this study, and 41 from blood donors from the 17 provinces and municipalities in China. Below the tree is a time scale from 1950 to 2015; The indications in this legend are the same as those in Fig. 3.

sequenced. Our results showed that 1b (61.7%) and 6a (20.2%) were the two most common subtypes in Guangzhou, followed in frequency by 2a (6.1%), 3b (6.1%), and 3a (5.2%). In

Fig. 5. Phylogeographic tree estimated with the partial Core-E1 sequences of subtype 6a isolates. A total of 155 subtype 6a sequences were used to generate the phylogeographic tree: 86 obtained in this study, 22 blood donors from Vietnam, and 47 from blood donors from the 17 provinces and municipalities in China. Below the tree is a time scale from 1960 to 2015. The indications in this legend are the same as those in Fig. 3.

addition, we identified subtype 4a and 6u in this region, accounting for HCV infection in only 1 (0.2%) and 2 (0.5%) individuals, respectively.

This study as well as others demonstrated that subtype 1b is the most prevalent subtype in Guangzhou, Guangdong Province, China [6-11, 13, 16]. However, the prevalence of subtype

1b decreased from 70% to 60% in recent years. The factors associated with the decreased subtype 1b prevalence in Guangdong Province are not clear. One plausible explanation is a shift in the main transmission route from blood transfusion and/or infusing of blood products after the introduction of routine screening blood donors with HCV tests to a route of intravenous drug use (IDU) and high-risk sexual behavior (HRSB) [17-22]. Furthermore, the decreased prevalence of subtype 1b may have impact on treatment outcomes, as 1b tends to be less likely responsive to interferon-based antiviral treatment than genotypes 2 and 3 [23-29]. HCV subtype 2a is the second most predominant subtype in China, accounting for 20% of HCV infected cases nationwide. However, 6a was found to be the second most predominant subtype with a 20.2% prevalence in Guangzhou, while 2a represents the third most predominant subtype (6.1%) in this study.

In an early analysis of HCV RNA sequences from 411 volunteer blood donors from 17 provinces and municipalities, which included 66 Guangzhou isolates [16], five phylogeographic trees consisting of HCV subtypes 1b, 2a, 3a, 3b, and 6a were constructed. However, Guangzhou has been a robust engine in driving fast economic development for decades in China and has now become a “World Production Center”, which undoubtedly influenced the epidemiological patterns of infectious diseases both locally and nationally [17]. Therefore, we employed the above-mentioned dataset as well as a large number of additional sequences from Guangzhou and Vietnam to construct five time-scaled phylogenetic trees with the Bayesian phylogeographic inference framework in the present study.

The tree for 1b shows six groups characteristic of different geographic distribution patterns and migration trends. Except for group D and E, which contain sequences with a substantial mixture of geographic origins, indicating the nationwide prevalence of group D and E isolates. Each of the other four groups contained sequences mostly from a single region. Group A was more prevalent in the northwest and frequently spread to other regions, while group B, C and F were more common in Guangzhou and occasionally appeared outside that region.

Interestingly, as a result of including a large number of Guangzhou sequences, the phylogeographic tree showed a more clear migration trend: 1b strains in other regions of China may have been spread and descended from Guangzhou during the 1970s and 1980s. Among them, the north-northeast and northwest became the second source regions to disseminate 1b strains to other regions. Factors driving the 1b strains to migrate to other regions from Guangzhou may include the close socio-economic links between Guangzhou and other regions in China, since this city serves as an important national transportation hub in China. Moreover, Guangzhou was the first region in China to undergo economic reform since 1978, resulting in profound social-economic changes and an influx of millions of immigrant workers. During holidays and busy farming seasons, those migrants frequently travel back and forth between Guangdong and their hometowns, facilitating pathogens spreading.

The phylogeographic tree for subtype 2a showed two statistically well supported groups. The smaller one had the majority of sequences originating in Guangzhou, while the larger one displayed isolates with a mixture of geographic origins but mainly from the northwest. The latter was in accordance with Lu's study, indicating that 2a strains in China may have originated in the northwest. Besides, 2a is considered exotic and is suggested to be transmitted to the Northwest China from Afghanistan, a known narcotic production center, via a drug trafficking route. Several land drug trafficking routes have been indicated to run through Afghanistan to China.

The phylogeographic tree for subtype 3a showed three clear separate geographic groups: one had origins almost exclusively in the northwest, one had origins almost exclusively in the southwest, and the third one had origins almost exclusively in Guangzhou. This pattern supports the hypothesis that subtype 3a could have been introduced into China from different neighboring countries.

In the 3b trees, two subsets were roughly divided. In contrast to the tree for subtype 3a, the phylogeographic tree for subtype 3b showed sequences with a mixture of geographic origins, to a certain extent. This indicates a migration trend from the southwest to Guangzhou,

and then spreading to the northwest, with a few migrations to the north-northeast and the southeast. According to Table 1, we speculate that the migration from the southwest to other regions was primarily by transmission via the IDU network, since 65.7% of the genotype 3b patients had a history of intravenous drug use. This is consistent with the known drug trafficking routes in Yunnan Province that link the Golden Triangle in Southeast Asian countries to China.

The phylogeographic tree for subtype 6a appears to show a trend of migration from Vietnam to China, which is consistent with the previous studies, though we have not included all core-E1 sequences from other countries especially Southeast Asia, owing to the lack of relevant data on GenBank. Besides, there appear to be a trend of 6a migration first to Guangzhou, then to many other regions in the present study, as only limited subtype 6a sequences of the partial Core-E1 region in Yunnan and Guangxi are available in Genbank. We believe the IDU network could have played a critical role in the introduction of 6a from Vietnam to Guangzhou, as 39.5% of genotype 6a patients had a history of intravenous drug use, which was much more common than other transmission modes. Furthermore, Guangzhou receives millions of migrant workers and visitors from across the country each year. Those frequently flowing people can carry HCV 6a from Guangdong Province back to their hometowns in different regions of China [30].

There are some limitations in our study. First, the data were biased toward Guangzhou, which may have impact on the root estimates in the phylogeographic tree. To relieve this concern, we used SPSS 17.0 software to generate random numbers for selecting samples from 1b and 6a sequences, and found that migration trends of GT1b and GT6a in China have little changes (unpublished data). Second, we made exhaustive searches in the GenBank nucleotide sequence database and download as many matched 1b and 6a sequences as possible. However, most of the matched sequences had to be eliminated from analysis because of a lack of important information such as ancestral geographical states and sampling time, resulting in different numbers of sequences among different regions.

In conclusion, our study suggests that there are seven HCV subtypes, 1b, 2a, 3a, 3b, 4a, 6a, and 6u present in Guangzhou, Guangdong Province, with subtype 1b dominant, followed by 6a, 2a, and 3b. Interestingly, though the distribution of HCV genotypes in Guangzhou is complex, our analyses do show the unique features. We suggest that HCV strains in Guangzhou may have a variety of geographic origins. In addition, genotype 6a has become endemic in Guangzhou, which led to spreading to many other regions of China. However, because of the possible data bias embedded this study, our conclusions need to be verified in new studies with large random sampled sequences.

Acknowledgements

This study was supported by the grants from the National Natural Sciences Foundation of China (81470856 and 31470263) and Science and Technology Development Fund of Shenzhen (No. JCYJ20120831144704365). The funding agencies had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Disclosure Statement

The authors declare that there are no conflicts of interest.

References

- 1 Easl recommendations on treatment of hepatitis c 2016. *J Hepatol* 2017;66:153-194.
- 2 Mohd HK, Groeger J, Flaxman AD, Wiersma ST: Global epidemiology of hepatitis c virus infection: New estimates of age-specific antibody to hcv seroprevalence. *Hepatology* 2013;57:1333-1342.

- 3 Fu Y, Wang Y, Xia W, Pybus OG, Qin W, Lu L, Nelson K: New trends of hcv infection in china revealed by genetic analysis of viral sequences determined from first-time volunteer blood donors. *J Viral Hepat* 2011;18:42-52.
- 4 Hepatitis c guidance: Aasld-idsa recommendations for testing, managing, and treating adults infected with hepatitis c virus. *Hepatology* 2015;62:932-954.
- 5 Smith DB, Bukh J, Kuiken C, Muerhoff AS, Rice CM, Stapleton JT, Simmonds P: Expanded classification of hepatitis c virus into 7 genotypes and 67 subtypes: Updated criteria and genotype assignment web resource. *Hepatology* 2014;59:318-327.
- 6 Rao H, Wei L, Lopez-Talavera JC, Shang J, Chen H, Li J, Xie Q, Gao Z, Wang L, Wei J, Jiang J, Sun Y, Yang R, Li H, Zhang H, Gong Z, Zhang L, Zhao L, Dou X, Niu J, You H, Chen Z, Ning Q, Gong G, Wu S, Ji W, Mao Q, Tang H, Li S, Wei S, Sun J, Jiang J, Lu L, Jia J, Zhuang H: Distribution and clinical correlates of viral and host genotypes in chinese patients with chronic hepatitis c virus infection. *J Gastroenterol Hepatol* 2014;29:545-553.
- 7 Pybus OG, Barnes E, Taggart R, Lemey P, Markov PV, Rasachak B, Syhavong B, Phetsouvanah R, Sheridan I, Humphreys IS, Lu L, Newton PN, Klennerman P: Genetic history of hepatitis c virus in east asia. *J Virol* 2009;83:1071-1082.
- 8 An Y, Wu T, Wang M, Lu L, Li C, Zhou Y, Fu Y, Chen G: Conservation in china of a novel group of hcv variants dating to six centuries ago. *Virology* 2014;464-465:21-25.
- 9 Wu T, Xiong L, Wang F, Xu X, Wang J, Lin F, Li C, Lu L, Zhou Y: A unique pattern of hcv genotype distribution on hainan island in china revealed by evolutionary analysis. *Cell Physiol Biochem* 2016;39:316-330.
- 10 Fu Y, Qin W, Cao H, Xu R, Tan Y, Lu T, Wang H, Tong W, Rong X, Li G, Yuan M, Li C, Abe K, Lu L, Chen G: Hcv 6a prevalence in guangdong province had the origin from vietnam and recent dissemination to other regions of china: Phylogeographic analyses. *PLoS One* 2012;7:e28006.
- 11 Tao J, Liang J, Zhang H, Pei L, Qian HZ, Chambers MC, Jiang Y, Xiao Y: The molecular epidemiological study of hcv subtypes among intravenous drug users and non-injection drug users in china. *PLoS One* 2015;10:e0140263.
- 12 Pybus OG, Rambaut A: Evolutionary analysis of the dynamics of viral infectious disease. *Nat Rev Genet* 2009;10:540-550.
- 13 Peng J, Lu Y, Liu W, Zhu Y, Yan X, Xu J, Wang X, Wang Y, Liu W, Sun Z: Genotype distribution and molecular epidemiology of hepatitis c virus in hubei, central china. *PLoS One* 2015;10:e0137059.
- 14 Slatkin M, Maddison WP: A clastic measure of gene flow inferred from the phylogenies of alleles. *Genetics* 1989;123:603-613.
- 15 Parker J, Rambaut A, Pybus OG: Correlating viral phenotypes with phylogeny: Accounting for phylogenetic uncertainty. *Infect Genet Evol* 2008;8:239-246.
- 16 Lu L, Wang M, Xia W, Tian L, Xu R, Li C, Wang J, Rong X, Xiong H, Huang K, Huang J, Nakano T, Bennett P, Zhang Y, Zhang L, Fu Y: 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;88:7120-7129.
- 17 Yan Z, Fan K, Wang Y, Fan Y, Tan Z, Deng G: Changing pattern of clinical epidemiology on hepatitis c virus infection in southwest china. *Hepat Mon* 2012;12:196-204.
- 18 Wang Y, Okamoto H, Mishiro S: Hcv genotypes in china. *Lancet* 1992;339:1168.
- 19 Zhang L, Zhang D, Chen W, Zou X, Ling L: High prevalence of hiv, hcv and tuberculosis and associated risk behaviours among new entrants of methadone maintenance treatment clinics in guangdong province, china. *PLoS One* 2013;8:e76931.
- 20 Zhou YB, Wang QX, Liang S, Gong YH, Yang MX, Nie SJ, Nan L, Yang AH, Liao Q, Yang Y, Song XX, Jiang QW: Hiv-, hcv-, and co-infections and associated risk factors among drug users in southwestern china: A township-level ecological study incorporating spatial regression. *PLoS One* 2014;9:e93157.
- 21 Piao HX, Yang AT, Sun YM, Kong YY, Wu XN, Zhang YZ, Ding B, Wang BE, Jia JD, You H: Increasing newly diagnosed rate and changing risk factors of hcv in yanjian prefecture, a high endemic area in china. *PLoS One* 2014;9:e86190.
- 22 Wang L, Tang W, Wang L, Qian S, Li YG, Xing J, Li D, Ding Z, Babu GR, Wang N: The hiv, syphilis, and hcv epidemics among female sex workers in china: Results from a serial cross-sectional study between 2008 and 2012. *Clin Infect Dis* 2014;59:e1-9.
- 23 Ebata M, Fukuda Y, Koyama Y, Hayakawa T, Kumada T, Nakano S: [hcv genotype as a predictor of response to ifn therapy in chronic hepatitis c]. *Nihon Rinsho* 1995;53:S954-958.

- 24 Asselah T: Daclatasvir plus sofosbuvir for hcv infection: An oral combination therapy with high antiviral efficacy. *J Hepatol* 2014;61:435-438.
- 25 Forns X, Gordon SC, Zuckerman E, Lawitz E, Calleja JL, Hofer H, Gilbert C, Palcza J, Howe AY, DiNubile MJ, Robertson MN, Wahl J, Barr E, Buti M: Grazoprevir and elbasvir plus ribavirin for chronic hcv genotype-1 infection after failure of combination therapy containing a direct-acting antiviral agent. *J Hepatol* 2015;63:564-572.
- 26 Ansaldi F, Orsi A, Sticchi L, Bruzzzone B, Icardi G: Hepatitis c virus in the new era: Perspectives in epidemiology, prevention, diagnostics and predictors of response to therapy. *World J Gastroenterol* 2014;20:9633-9652.
- 27 Fried MW, Shiffman ML, Reddy KR, Smith C, Marinos G, Goncales FJ, Haussinger D, Diago M, Carosi G, Dhumeaux D, Craxi A, Lin A, Hoffman J, Yu J: Peginterferon alfa-2a plus ribavirin for chronic hepatitis c virus infection. *N Engl J Med* 2002;347:975-982.
- 28 Song ZZ: Randomized study of danoprevir/ritonavir-based therapy for hcv genotype 1 patients with prior partial or null responses to peginterferon/ribavirin. *J Hepatol* 2015;63:769-770.
- 29 Layden-Almer JE, Ribeiro RM, Wiley T, Perelson AS, Layden TJ: Viral dynamics and response differences in hcv-infected african american and white patients treated with ifn and ribavirin. *Hepatology* 2003;37:1343-1350.
- 30 Rong X, Xu R, Xiong H, Wang M, Huang K, Chen Q, Li C, Liao Q, Huang J, Xia W, Luo G, Ye X, Zhang M, Fu Y: Increased prevalence of hepatitis c virus subtype 6a in china: A comparison between 2004-2007 and 2008-2011. *Arch Virol* 2014;159:3231-3237.