Received Date : 01-Oct-2016
Revised Date : 30-Nov-2016
Accepted Date : 19-Dec-2016
Article type : Original Paper

Population dynamics of HCV subtypes in injecting drug-users on methadone maintenance treatment in China associated with economic and health reform

Running title: Population dynamics of HCV subtypes in IDUs on MMT

Sheng Zhou[1, 2#], Eleonora Cella[3,4,5#], Wang Zhou[2#*], Wen-Hua Kong[2], Man-Qing Liu[2], Pu-Lin Liu[2], Massimo Ciccozzi[3,7], Marco Salemi[5,6], Xin-Guang Chen[2,8*]

[1] Department of Epidemiology, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD, USA

[2] Wuhan Centers for Disease Prevention and Control, Wuhan, China

[3] Department of Infectious Parasitic and Immunomediated Diseases, Reference Centre on Phylogeny, Molecular Epidemiology and Microbial Evolution (FEMEM)/Epidemiology Unit, Istituto Superiore di Sanità, Rome, Italy

[4] Public Health and Infectious Diseases, Sapienza University, Rome, Italy

[5] Department of Pathology, Immunology, and Laboratory Sciences, College of Medicine, University of Florida, Gainesville, FL, USA

[6] Emerging Pathogens Institute, University of Florida, Gainesville, FL, USA

[7] University Hospital Campus Bio-Medico, Rome, Italy

[8] Department of Epidemiology, College of Public Health and Health Profession & College of Medicine, University of Florida, Gainesville, FL, USA

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1111/jvh.12677

[#] **These authors contributed equally to this work.**

***Corresponding authors:**

Dr. Wang Zhou, Wuhan Centers for Disease Prevention and Control, 24 Jianghan N. Rd, Wuhan 430015, China. Tel: +86-27-85801628; Fax: +86-27-85805227. Email: rising-up@hotmail.com

Prof. Xin-Guang Chen, Department of Epidemiology, College of Public Health and Health Profession & College of Medicine, University of Florida, 2004 Mowry Road, PO Box 100231, Gainesville, FL 32610-0231, USA. Tel: +1 352-273-5468. Email: jimax.chen@ufl.edu

ABSTRACT

The extensive genetic heterogeneity of hepatitis C virus (HCV) requires in-depth understanding of the population dynamics of different viral subtypes for more effective control of epidemic outbreaks. We analyzed HCV sequences data from 125 participants in Wuhan, China. These participants were newly infected by subtype 1b (n=13), 3a (n=15), 3b (n=50), and 6a (n=39) while on methadone maintenance treatment (MMT). Bayesian phylogenies and demographic histories were inferred for these subtypes. Participants infected with HCV-1b and 3a were clustered in well-supported monophyletic clades, indicating local sub-epidemics. Subtypes 3b and 6a strains were intermixed with other Chinese isolates, as well as isolates from other Asian countries, reflecting ongoing across geographic boundary transmissions. Subtypes 1b and 3a declined continuously during the past ten years, consistent with the health and economic reform in China, while subtype 3b showed ongoing exponential growth and 6a was characterized by several epidemic waves, possibly related to the recently growing number of travelers between China and other Asian countries. In conclusion, results of this study suggest that HCV subtype 3b and 6a sub-

epidemics in China are currently not under control, and new epidemic waves may emerge given the rapid increase in international traveling following substantial economic growth.

Importance

By analysis of the recent sequence data among drug users in Wuhan who were under methadone maintenance treatment, several HCV subtypes and transmission clusters were detected. The population dynamics of the detected subtypes were consistent with the economic growth and health care reform in China over the past 30 years, while the detected clusters provided evidence regarding the significance of continued injecting drug use and sexual risk behavior in fueling the epidemic. These findings provided new evidence regarding HCV in Wuhan and suggest urgent need for more effective prevention interventions targeting specific risk behaviors and new domestic and international collaborative prevention strategies for potential new waves of different HCV subtype epidemics.

Key words: HCV subtypes; Phylogenetics; HCV demographic history; Wuhan City; China

INTRODUCTION

Hepatitis C virus (HCV) is a single, positive-stranded RNA virus from the *Flaviviridae* family. Its genome consists of approximately 9600 base pairs, characterized by high levels of genetic variability(1, 2). HCV infection is the leading risk factor for chronic hepatitis, liver cirrhosis and hepatocellular carcinoma; it affects 130-150 million people worldwide(3). HCV infection is a global epidemic with Central and East Asia and North Africa being the focal regions. Injecting drug users (IDUs) is one of the at-risk populations for HCV infection.

HCV strains are classified into seven genotypes (1-7) based on phylogenetic and sequence analyses of the whole viral genome. Within these genotypes, HCV is further classified into 67 confirmed and 20 provisional subtypes(4). Genotypes 1-3 circulate worldwide with subtypes 1a and 1b accounting for 60% of the total infection(3). In addition, genotypes 1-3 are the most common subtypes circulating in Europe, North America, and Japan; genotype 2 is less frequently represented and is most prevalent in South America, Southeast Asia and Western Africa(4); and genotype 3 is an endemic strain spreading within Southeast Asia with variable distributions across different countries in the region. Genotype 4 is found primarily in the Middle East, Egypt, and central Africa; genotype 5 spreading is almost exclusively confined within South Africa; and genotypes 6 and 7 are prevalent in Asia(3).

HCV genotypes 1 and 2 (69.6% and 19.5%, respectively) are the two most prevalent HCV infections in China.(5) Within genotype 1, subtype 1b (68.4%) is more prevalent. Other subtypes are less prevalent, including 3a, 3b and 6a, and they are found primarily in the southern provinces of China. HCV subtype 1b is more likely to be transmitted through blood transfusion and other medical procedures, while subtype 6a is more likely to spread by injecting drug use and/or unprotected sex(6). Relative to genotype 1, infections with genotype 6 are more likely to be detected in younger patients(7). In a previous study with follow-up data(8), a high seroconversion rate (46.3%) of HCV was observed among a sample of participants who were registered for methadone maintenance treatment (MMT) in the Wuhan, a provincial capital city located in central China. The results suggest one or more relatively recent epidemic outbreaks during 2006-2011.

To curtail the HCV epidemic, there is an urgent need to advance our understanding of the viral population dynamics of different viral subtypes. Such data are essential to support evidence-based public health decision-making and effective prevention intervention. In this study, HCV sequence data were obtained from participants enrolled in our research project to investigate the molecular and demographic history of various HCV subtypes circulating in Wuhan, as well as their associations with risk behaviors and changes in public health policy.

MATERIALS AND METHODS

Participants and survey data

We analyzed HCV sequence data from 125 participants who converted to HCV positive while receiving MMT through a public health agency in Wuhan, China. Participants were identified through medical record review and the procedure of survey data collection was described in detail elsewhere(8). In brief, medical records for a total of 16085 methadone users enrolled during 2001-2006 for MMT were reviewed, of whom 12755 received HCV tests at entrance. Among those who were tested HCV^{-ve} at baseline, 1700 agreed to participate in the HCV study sponsored by the Wuhan Center of Disease for Disease Prevention and Control (CDC). Among the HCV^{-ve} patients, 555 eventually became HCV^{+ve} while on MMT, of whom 206 remained on MMT at the time when this study was conducted. Among these patients, 178 agreed to participate in the study, and 125 were confirmed to have high viral load (1.0×10^3 IU/ml < HCV RNA < 1.0×10^{10} IU/ml in real time PCR). Data regarding demographic factors, injection drug use (if sharing needles), and unprotected sex (not always use a condom during sex) were collected through in-person interviews conducted in 2012 and described by Zhou and colleagues(8).

The study protocol was approved by the Institutional Review Board at Wuhan CDC, and written informed consent was obtained from all participants. All experiments were conducted in accordance with the approved study protocol.

Blood samples

Blood specimens (10ml whole blood) were collected with K3-EDTA tubes. The collected samples were immediately placedt at +4°C, and then transferred to a designated laboratory at Wuhan CDC on the same day. The extracted plasma samples were stored at -20°C in aliquots for analysis. All specimens were re-tested for anti-HCV antibodies using a third generation enzyme-linked immunoassay (EIA-3) (Kehua Biotechnology Inc., Shanghai, China). The assay was performed according to the standard test procedures described by the manufacturer. Specimens found to be initially reactive by EIA-3 were repeated in duplicate. Repeat reactive specimens were used to conduct further HCV RNA analyses.

HCV RNA analysis

HCV RNA was extracted from the blood specimens with the HCV RT-PCR ASSAY V2 kit (QIAGEN, Düsseldorf, Germany). The concentration of HCV RNA was measured by real-time polymerase chain reaction (RT-PCR) in a 7300 Real Time PCR system. The isolated RNA was amplified with the conventional RT-PCR with two sets of primers (Ac2, Sc2 and S7, A7) targeting the 5'-UTR and Core (335 bp) regions. The amplified DNA products were analyzed by agarose gel electrophoresis (2%) containing 0.5μg/mL ethidium bromide, followed by visualization under ultraviolet light. Sequencing and subtyping of the HCV genes was conducted through contract with Sangon Biotech Co., Ltd. (Shanghai, China). Sequence alignment was cperformed using the CLUSTAL X program (version 2.1, University College

Dublin, Ireland). The sequences obtained were submitted to GenBank with the following accession numbers: JX944384 - JX944467, KC348398 - KC348438.

Sequence Data set

Subtype-specific alignments of the four most frequently observed subtypes revealed that 117 out of 125 sequences belonged to subtype 1b (13 sequences), 3a (15), 3b (50) and 6a (39), respectively using the Clustal algorithm (9) followed by manual editing, using Bioedit (10). Each alignment also included reference sequences downloaded from NCBI (http://www.ncbi.nlm.nih.gov), collected during outbreaks in drug user populations since 1988 (193 HCV-1b sequences isolated during 1988-2011; 79 HCV-3a sampled during 2000-2012; 82 HCV-3b sampled during 1995-2012, and 85 HCV-6a sampled between 1995-2012). Inclusion criteria for the reference sequences were: (1) Data were published in peer-review journals; (2) no uncertainty regarding the subtype assignment; (3) detailed date and place or origin (city/state) were clearly established and reported in the original publication. Due t0 over-representation of HCV sequences from some countries, country-specific sequences were randomly selected for each subtype-specific data set to ensure similar weight for individual sampling locations. The number of sequences included in the final datasets was 17 for 1b, 13 for 3a, 62 for 3b and 72 for 6a, respectively.

Likelihood mapping analysis

Phylogenetic signal in each dataset of aligned DNA or amino acid sequences was investigated with the likelihood mapping method with TREE-PUZZLE (11). In brief, the three likelihoods for the three unrooted tree topologies of every possible group of four sequences (quartet) was estimated and simultaneously displayed as a dot in an equilateral triangle (a

likelihood map) where the distance between the dot and each side of the triangle is proportional to the likelihood of a specific topology. In such a likelihood map, three areas can be distinguished: (a) the three corners, representing fully resolved tree topologies, i.e. the presence of tree like phylogenetic signal in the data; (b) the center, representing the star-like phylogeny, and (c) the three side areas, indicating network-like phylogeny, i.e. presence of recombination or conflicting phylogenetic signals. Findings from extensive simulation studies and analysis of real data sets suggest that >30% dots in the central area indicate significant phylogenetic noise (star-like signal), reducing the reliability of phylogeny inference(11-13).

Phylogenetic analysis

For each data set, the best fitting nucleotide substitution model was selected with the results of the hierarchical likelihood ratio test implemented with the Model test software version 3.7 (14). Maximum likelihood (ML) phylogenetic trees were then constructed using the selected model with PhyML3.0 (http://www.atgc-montpellier.fr/phyml/). Statistical robustness and reliability of the branching order within the phylogenies were assessed by bootstrapping analysis (1000 replicates), by the fast likelihood-based method (approximate likelihood-ratio test), as well as the parametric Bayesian-like transformation of aLRT (aBayes and Chi2-based), and nonparametric Shimodaira-Hasegawa-like procedure (SH-aLRT) (15).

Bayesian coalescence inference

The molecular clock hypothesis for individual HCV data sets was investigated by calibrating either a strict (with a lognormal prior for the clock rate) or a relaxed (with a lognormal prior for the ucld mean) molecular clock, using a constant coalescent prior and dated tips

corresponding to the sequence sampling times, with the Bayesian Markov Chain Monte Carlo (MCMC) method implemented in BEAST v1.8 (http://beast.bio.ed.ac.uk)(16, 17). In order to investigate the demographic history of a HCV subtype, four independent MCMC runs were also carried out enforcing the best fitting clock model (relaxed molecular clock, see Results) and one of the following coalescent priors: parametric constant population size and exponential growth, and non-parametric smooth skyride plot Gaussian Markov Random Field (GMRF)(16, 18, 19) and Bayesian Skyline plot (BSP).(16, 17) Marginal likelihood estimates for each demographic model were obtained using path sampling and stepping stone analyses(20-22). Uncertainty in the estimates was indicated by 95% highest posterior density (95% HPD) intervals, and the best fitting model for each data set was established by calculating the Bayes Factors (BF)(21, 23). In practice, any two models can be compared to evaluate the strength of evidence against the null hypothesis (H_0), defined as the one with the lower marginal likelihood: 2lnBF < 2 indicates no evidence against H_0; 2–6, weak evidence; 6–10: strong evidence, and > 10 very strong evidence. For each data set, the MCMC sampler was run for at least 50 x 10^6 generations, sampling every 5000 generations. Proper mixing of the MCMC was assessed by calculating the effective sample size (ESS) of each parameter. Only parameter estimates with ESS > 200 were accepted. Phylogeographic analysis was conducted on the four datasets, by using the continuous time Markov Chain (CTMC) process over discrete sampling locations implemented in BEAST(24) with the Bayesian Stochastic Search Variable Selection (BSSVS) model, which allows diffusion rates to be zero with a positive prior probability. Maximum clade credibility (MCC) trees were selected from the posterior tree distribution after a 10% burn-in using the Tree Annotator program version 1.8 included in the BEAST package(16, 17). Statistical support for specific monophyletic clades was assessed by calculating the posterior probability. All xml files used for the Bayesian analyses are available from the authors upon request.

RESULTS

Demographic and epidemiological characteristics of the study participants

The participants were distributed in four districts of Wuhan, including Jiangan (n=30), Qiaokou (n=12), Qingshan (n=5) and Wuchang (n=70) (Figure 1). The majority (85, 72.65 %) were male; the median age [inter-quarter range IQR] was 40 [33.7, 45.0] years, and the median [IQR] age at which a drug was used for the first time was 29 [24.0, 35.2]. Most of the participants (90, 76.9%) were injection drug users (IDUs) and a large number of participants (42, 36%) reported no consistent condom use (not always use a condom during sex in the past 3 months). The median viral load [IQR] was 9610000 [184000- 4390000].

Phylogenetic and spatial distribution patterns of HCV subtypes in Wuhan

Despite the short length of the HCV genomic region for phylogenetic inference (330 base pairs of the 5UTR-Core, see Methods), results from the likelihood mapping analysis showed low levels of phylogenetic noise (12.4% to 21.6 %) in all sequence data (Supplementary figure S1 panel a-d). This finding indicated sufficient signal for phylogeny inference and was further confirmed by the proportion of parsimony informative sites, ranging from 13.3% (subtype 6a data set) to 25.5%(subtype 3b data set). For each of the four detected HCV subtypes, the molecular clock hypothesis was tested by Bayes factor's comparison of the strict (with a lognormal prior for the clock rate) *vs.* the relaxed (with a lognormal prior for the ucld mean) clock model, with the evolutionary rate in each model estimated from dated tips corresponding to the known sampling time of the sequences. As expected, the relaxed clock consistently showed a better fit than the strict clock model (Supplementary table S1). The demographic history of each viral subtype was inferred under a relaxed molecular clock by using both parametric models with constant population size and exponential growth, and

non-parametric models with a smooth skyride plot Gaussian Markov Random Field (GMRF) and Bayesian skyline plot (BSP). The BSP model outperformed the other models for subtypes 1b, 3a and 6a, while the exponential growth model was the best fit for subtype 3b (Supplementary table S2). Evolutionary rate estimates (using the best fitting demographic model) ranged, as expected, between 2.3 10^{-3} (subtype 1b) to 1.1 10^{-2} (subtype 6a) nucleotide substitutions per site per year with overlapping 95% high density intervals (95% HPD) among subtypes (Supplementary Table S3).

For each of the data sets used in this analysis, similar topologies were observed from both maximum likelihood (ML) trees and Bayesian trees (data not shown). All but two HCV subtype 1b strains from Wuhan clustered within a highly supported monophyletic clade (Figure 2) that also included several Chinese reference sequences. The Wuhan clade shared a most recent common ancestor with a Thailand lineage, suggesting an introduction from Thailand to China between the mid-1950s and the mid-1960s. Two isolates, one from Jiangan District and another from Wuchang District, formed a cluster with US reference sequences outside the major Wuhan clade. Within the Wuhan clade, three well-supported sub-clusters of closely related sequences, possibly indicating epidemiologically linked cases, were evident (Figure 2). Sub-cluster A included two sequences of IDUs from Jiangan and Wuchang Districts, who reported no consistent condom use during sex; sub-cluster B included two sequences of IDUs from Wuchang District who reported no consistent condom use, while sub-cluster C included two IDUs both from Qingshan District and reported no consistent condom use.

Overall, the data are consistent with a regional HCV epidemic in the Wuhan districts, originated from a single or limited number of introductions and characterized by putative transmission clusters possibly fostered by injecting drug use and/or unprotected sexual practice.

A similar pattern was observed for HCV subtype 3a strains. Data from the MCC tree indicated that Wuhan sequences clustered within a highly supported monophyletic clade (Figure 3), with exception of one sequence from Jiangan District and another from Wuchang District. Outside of the Wuhan clade, there was a Chinese reference sequence branching off to cluster with a Venezuelan reference strain. The Wuhan clade of HCV-3a strain, like the 1b subtype strain, shared a common ancestor with a Thailand lineage. The origin appeared to be much more recent, consistent with an introduction between 2000 and 2005 (Figure 3). Two well-supported sub-clades within the strain of HCV-3a Wuhan clade could also be distinguished: one including sequences of non-IDU participants from Wuchang District and another including sequences of IDU participants from Jiangan District. Both participants reported no consistent condom use.

A very different pattern was observed for HCV subtypes 3b and 6a. Subtype 3b sequences from Wuhan were highly intermixed with both Chinese and non-Chinese references, including strains from Bangladesh, Canada, India, Cambodia, Myanmar, Pakistan, Thailand, and Vietnam (Figure 4). Within the tree, several sub-clades (A-F) of closely related sequences representing putative transmission chains were statistically supported, all originating after 2010. In clade A, two Chinese reference sequences clustered with two IDU and two non-IDU sequences from the Jiangan, Qiaokou and Wuchang Districts (Figure 4).

Clade B included three IDU sequences from Wuchang participants who reported no consistent condom use. Clade C included two IDU sequences from Jiangan with either no or frequent use of condom. Clade D and F included two sequences of two IDUs from Wuchang reporting occasional condom use. Clade E included non-IDUs from Wuchang who reported no consistent condom use.

Overall, the pattern was consistent with multiple separate introductions of HCV-3a into Wuhan, either from other Chinese regions or from abroad, and the discrete transmission chains potentially originating from injecting drug use (needle sharing) and/or unprotected sex.

Similarly, HCV 6a subtype phylogeny showed that Wuhan strains intermixed with both Chinese and non-Chinese (India, Japan, Vietnam, Iran) references. One major clade, including most of the strains from Wuhan samples (80%) intermixed with strains from China, Iran and Japan, was statistically well supported, while the remaining Wuhan strains formed a monophyletic clade branching off Vietnamese sequences. Within the major clade, three well-supported sub-clades of sequences from Wuhan of possible epidemiologically linked cases, again of relatively recent origin (2008-2010), were also evident (Figure 5). Clade A included two IDUs sequences from Wuchang branching out from an IDU sequence from Qiaokou, which in turn branched out from a non-IDU sequence from Jiangan. Cluster B contained two sequences of IDUs from Wuchang who reported frequent condom use; Cluster C included both IDUs and non-IDUs from Wuchang.

HCV population dynamics of major subtypes circulating in Wuhan

Demographic histories inferred from individual subtype-specific alignments displayed two distinct patterns of effective viral population size (*Ne*) – a genetic diversity measure representing the number of genomes effectively contributing to the next generation over time (Figure 6). The estimated *Ne* for subtype 1b and 3a showed logistic-like growth that, after an exponential phase during the 1970-1990s for HCV-1b and the 1990s for HCV-3a, leveled off between 2000 and 2005, and was followed by a decline during the next five years. On the other hand, the estimated *Ne* for subtype 3b was characterized by an exponential growth during the whole observation period, suggesting an uncontrolled epidemic; while the *Ne* for subtype 6a BSP showed series of peaks and troughs, suggesting multiple epidemic waves probably with multiple introductions of HCV strains into Wuhan. Interestingly, *Ne* estimates for subtypes 3b and 6a from 2005 on are 1.5-2 logs lower than subtypes 1b and 3a (Figure 6).

DISCUSSION

HCV infection in China is a significant public health concern. In the present study, we analyzed HCV viral sequence data collected among a sample of 117 participants in Wuhan, China. These patients were newly diagnosed with HCV infection while on MMT. We investigated the spatial distribution and evolutionary history of four predominant HCV subtypes 1b 3a, 3b and 6a in the region. MMT patients infected with subtypes 1b and 3a were clustered in well-supported monophyletic clades, indicating that the local epidemics of these two subtypes probably originated from a limited number of introductions from Thailand during the 1950s-1960s for 1b and around 2000-2005, for 3a. On the other hand, HCV subtypes 3b and 6a strains were highly intermixed with isolates within China and

isolates from other Asian countries, suggesting potential cross-geographic boundary transmission.

Findings from the HCV phylogeography analysis also showed cross-district spreading within Wuhan. The four HCV subtypes circulating in Wuhan intermixed with sequences sampled in the three (Jiangan, Wuchang and Qiakou) of the four districts. Furthermore, these infections were epidemiologically linked with injecting drug use and/or unprotected sexual practice. This is consistent with findings from previous research that most patients on MMT continue to use illicit opioids and engage in unprotected sex and other risk behaviors, increasing their risk for HCV infection(25). Indeed, unsafe drug injections, defined as sharing and reuse of syringes or needles among patients without sterilization, results in 2.3-4.7 million HCV infections globally every year(26), and its impact has been recognized in many developing countries, including China(27-30).

In addition to temporal patterns, findings regarding the HCV demographic history showed two distinct trends. Subtypes 1b and 3a, predominantly spread within China are characterized by continuous declines in effective population size since 2005; while subtype 3b and 6a, primarily cross-border transmission between China and Southeast Asian countries, displayed an upward trend during the same period with subtype 6a showing several epidemic waves. In particular, findings of our study showed growing trends in subtype 1b since the 1980s and 3a since the early 1990s, both of which leveled off around the early 2000s and eventually declined. The time frame of this exponential HCV growth is consistent with the tremendous negative health consequences associated with the economic reforms and the marketization of medical health services in China, and the

subsequent initiatives of health care reform by the Chinese government since 2005 to emphasize prevention and primary care(31).

Contrary to these two domestic HCV subtypes, the evolutionary and demographic patterns for the two international subtypes 3b and 6a appear to be consistent with the growing number of travelers between China and other Asian countries (particularly the neighboring Asian countries) over the past two decades. It is worth noticing that when the epidemics of subtypes 1b and 3a were at their peak levels after a phase of exponential growth, *Ne* estimates for these two subtypes were substantially higher than subtypes 3b and 6a. This could be due to the longer period of the subtype 3a and 6b epidemics than those of 1b and 3a in Asia as reflected in our reconstructed temporal trends. However, considering the increasing trends in *Ne* for subtypes 3b and 6a inferred from 2005 on, it is likely that, if unchecked, these subtypes may increase within the next decade to the same level as subtypes 1b and 3a in the past. This finding stresses the urgent need for new measures to prevent the cross-Asian country HCV transmission.

In conclusion, the findings of this study suggest potential new waves and even HCV subtype 3b and 6a epidemics, given the rapid increase in international traveling following the rapid economic growth and the prevalence of drug use and unprotected sexual practices. Further research with more diverse samples is also needed to further confirm the results we observed in this study.

REFERENCES

1. **Simmonds P, Holmes EC, Cha TA, Chan SW, McOmish F, Irvine B, Beall E, Yap PL, Kolberg J, Urdea MS.** 1993. Classification of hepatitis C virus into six major genotypes and a series of subtypes by phylogenetic analysis of the NS-5 region. J Gen Virol **74 (Pt 11):**2391-2399.

2. **Simmonds P.** 2004. Genetic diversity and evolution of hepatitis C virus--15 years on. J Gen Virol **85:**3173-3188.

3. **Brostrom B, Granich R, Gupta S, Samb B.** 2014. Reimagining HIV Testing in an Era of ART. AIDS Res Hum Retroviruses **Suppl 1:**A87.

4. **Messina JP, Humphreys I, Flaxman A, Brown A, Cooke GS, Pybus OG, Barnes E.** 2015. Global distribution and prevalence of hepatitis C virus genotypes. Hepatology **61:**77-87.

5. **Jia L, Yu J, Yang J, Song H, Liu X, Wang Y, Xu Y, Zhang C, Zhong Y, Li Q.** 2009. HCV antibody response and genotype distribution in different areas and races of China. Int J Biol Sci **5:**421-427.

6. **Fu Y, Wang Y, Xia W, Pybus OG, Qin W, Lu L, Nelson K.** 2011. New trends of HCV infection in China revealed by genetic analysis of viral sequences determined from first-time volunteer blood donors. J Viral Hepat **18:**42-52.

7. **Chen YD, Liu MY, Yu WL, Li JQ, Peng M, Dai Q, Liu X, Zhou ZQ.** 2002. Hepatitis C virus infections and genotypes in China. Hepatobiliary Pancreat Dis Int **1:**194-201.

8. **Zhou W, Wang X, Zhou S, Xie N, Liu P, Luo L, Peng J, Liu M, Desrosiers A, Schottenfeld R, Chawarski MC.** 2015. Hepatitis C seroconversion in methadone maintenance treatment programs in Wuhan, China. Addiction **110:**796-802.

9. **Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG.** 1997. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res **25:**4876-4882.

10. **Hall TA.** 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symp **41:**95-98.

11. **Schmidt HA, Strimmer K, Vingron M, von Haeseler A.** 2002. TREE-PUZZLE: maximum likelihood phylogenetic analysis using quartets and parallel computing. Bioinformatics **18:**502-504.

12. **Azarian T, Lo Presti A, Giovanetti M, Cella E, Rife B, Lai A, Zehender G, Ciccozzi M, Salemi M.** 2015. Impact of spatial dispersion, evolution, and selection on Ebola Zaire Virus epidemic waves. Sci Rep **5:**10170.

13. **Lo Presti A, Cella E, Giovanetti M, Lai A, Angeletti S, Zehender G, Ciccozzi M.** 2015. Origin and evolution of Nipah virus. J Med Virol.

14. **Posada D, Buckley TR.** 2004. Model selection and model averaging in phylogenetics: advantages of akaike information criterion and bayesian approaches over likelihood ratio tests. Syst Biol **53:**793-808.

15. **Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O.** 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst Biol **59:**307-321.

16. **Drummond AJ, Rambaut A, Shapiro B, Pybus OG.** 2005. Bayesian coalescent inference of past population dynamics from molecular sequences. Mol Biol Evol **22:**1185-1192.

17. **Drummond AJ, Rambaut A.** 2007. BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evol Biol **7:**214.

18. **Drummond AJ, Nicholls GK, Rodrigo AG, Solomon W.** 2002. Estimating mutation parameters, population history and genealogy simultaneously from temporally spaced sequence data. Genetics **161:**1307-1320.

19. **Minin VN, Bloomquist EW, Suchard MA.** 2008. Smooth skyride through a rough skyline: Bayesian coalescent-based inference of population dynamics. Mol Biol Evol **25:**1459-1471.

20. **Baele G, Lemey P.** 2013. Bayesian evolutionary model testing in the phylogenomics era: matching model complexity with computational efficiency. Bioinformatics **29:**1970-1979.

21. **Baele G, Lemey P, Vansteelandt S.** 2013. Make the most of your samples: Bayes factor estimators for high-dimensional models of sequence evolution. BMC Bioinformatics **14:**85.

22. **Baele G, Li WL, Drummond AJ, Suchard MA, Lemey P.** 2013. Accurate model selection of relaxed molecular clocks in bayesian phylogenetics. Mol Biol Evol **30:**239-243.

23. **Kass RE, Raftery AE.** 1995. Bayes factors. Journal of the American statistical association **90:**773-795.

24. **Lemey P, Rambaut A, Drummond AJ, Suchard MA.** 2009. Bayesian phylogeography finds its roots. PLoS Comput Biol **5:**e1000520.

25. **Lin C, Wu Z, Rou K, Yin W, Wang C, Shoptaw S, Detels R.** 2010. Structural-level factors affecting implementation of the methadone maintenance therapy program in China. J Subst Abuse Treat **38:**119-127.

26. **Kane A, Lloyd J, Zaffran M, Simonsen L, Kane M.** 1999. Transmission of hepatitis B, hepatitis C and human immunodeficiency viruses through unsafe injections in the

developing world: model-based regional estimates. Bull World Health Organ **77:**801-807.

27. **Frank C, Mohamed MK, Strickland GT, Lavanchy D, Arthur RR, Magder LS, El Khoby T, Abdel-Wahab Y, Aly Ohn ES, Anwar W, Sallam I.** 2000. The role of parenteral antischistosomal therapy in the spread of hepatitis C virus in Egypt. Lancet **355:**887-891.

28. **Stark K, Poggensee G, Hohne M, Bienzle U, Kiwelu I, Schreier E.** 2000. Seroepidemiology of TT virus, GBC-C/HGV, and hepatitis viruses B, C, and E among women in a rural area of Tanzania. J Med Virol **62:**524-530.

29. **Chowdhury A, Santra A, Chaudhuri S, Dhali GK, Chaudhuri S, Maity SG, Naik TN, Bhattacharya SK, Mazumder DN.** 2003. Hepatitis C virus infection in the general population: a community-based study in West Bengal, India. Hepatology **37:**802-809.

30. **Khan UR, Janjua NZ, Akhtar S, Hatcher J.** 2008. Case-control study of risk factors associated with hepatitis C virus infection among pregnant women in hospitals of Karachi-Pakistan. Trop Med Int Health **13:**754-761.

31. **Chen X, Wang PG.** 2014. Social change and national health dynamics in China. Chinese Journal of Population Science **2:**63-73.

Figure legends

Figure 1. Map of Wuhan, China. The four different sampling locations (Jiangan, Qiaokou, Qingshan and Wuchang) are indicated. The map is adapted from the original Chinese map from the National Administration of Surveying, Mapping and Geoinformation of China (http://219.238.166.215/mcp/index.asp).

Figure 2. Bayesian phylogeny of HCV 1b sequences. (Left panel) maximum clade credibility (MCC) tree with branch lengths scaled in time, including HCV 1b sequences from Wuhan, China as well as reference strains from GenBank. Branches are coloured according to the geographic location of sampled sequences (tip branches) or location of ancestral lineages inferred by Bayesian phylogeography (internal branches) according to the colour legend in the figure. The three asterisks along the branch represent significant statistical support for the clade by Bayesian posterior probability $p = 1$, as well as significant support by fast likelihood-based SH-like, aBayes, Chi2-based tests and bootstrapping (≥ 90%) in the maximum likelihood tree. (Right panel) expansion of the monophyletic clade including the majority of Wuhan strains, highlighted in the MCC tree by a circle. Branch lengths are scaled in years according to the bar at the bottom of the tree. Self-reported condom use and the drug injection for each of the strains sampled from Wuhan patients are indicated by symbols according to the legend in the figure. One asterisk along the branches represents significant statistical support for the clade subtending that branch ($p \geq 0.9$). Recent clusters potentially representing epidemiologically linked cases are indicated by letters. The tree with indicating GenBank accession numbers for tip labels is given in Supplementary Figure S2.

Figure 3. Bayesian phylogeny of HCV 3a sequences. (Left panel) maximum clade credibility tree (MCC) with branch lengths scaled in time, including HCV 3a sequences from Wuhan, China as well as reference strains from GenBank. Branches are coloured according to the geographic location of sampled sequences (tip branches) or location of ancestral lineages inferred by Bayesian phylogeography (internal branches) according to the colour legend in the figure. The two asterisks along the branch represent significant support by fast likelihood-based SH-like, aBayes, Chi2-based tests and bootstrapping (≥ 90%) in the maximum likelihood tree. (Right panel) expansion of the monophyletic clade including the

majority of Wuhan strains, highlighted in the MCC tree by a circle. Branch lengths are scaled in years according to the bar at the bottom of the tree. Self-reported condom use and the drug injection for each of the strains sampled from Wuhan patients are indicated by symbols according to the legend in the figure. One asterisk along the branches represents significant statistical support for the clade subtending that branch ($p \geq 0.9$). Recent clusters potentially representing epidemiologically linked cases are indicated by letters. The tree with indicating GenBank accession numbers for tip labels is given in Supplementary Figure S3.

Figure 4. Bayesian phylogeny of HCV 3b sequences. (Left panel) maximum clade credibility tree (MCC) with branch lengths scaled in time, including HCV 3b sequences from Wuhan, China as well as reference strains from GenBank. Branches are coloured according to the geographic location of sampled sequences (tip branches) or location of ancestral lineages inferred by Bayesian phylogeography (internal branches) according to the colour legend in the figure. The three asterisks along the branch represent significant statistical support for the clade by Bayesian posterior probability $p \geq 0.9$. (Right panel) expansion of the monophyletic clade including most of Wuhan strains, highlighted in the MCC tree by a circle. Branch lengths are scaled in years according to the bar at the bottom of the tree. Self-reported condom use and the drug injection for each of the strains sampled from Wuhan patients are indicated by symbols according to the legend in the figure. One asterisk along the branches represents significant statistical support for the clade subtending that branch ($p \geq 0.9$). Recent clusters potentially representing epidemiologically linked cases are indicated by letters. The tree with indicating GenBank accession numbers for tip labels is given in Supplementary Figure S4.

Figure 5. Bayesian phylogeny of HCV 6a sequences. (Left panel) maximum clade credibility tree (MCC) with branch lengths scaled in time, including HCV 6a sequences from Wuhan, China as well as reference strains from GenBank. Branches are coloured according to the geographic location of sampled sequences (tip branches) or location of ancestral lineages inferred by Bayesian phylogeography (internal branches) according to the colour legend in the figure. The asterisk along the branch represents significant statistical support for the clade by Bayesian posterior probability $p \geq 0.9$. (Right panel) expansion of the monophyletic clade including most of Wuhan strains, highlighted in the MCC tree by a circle. Branch lengths are scaled in years according to the bar at the bottom of the tree. Self-reported condom use and the drug injection for each of the strains sampled from Wuhan patients are indicated by symbols according to the legend in the figure. One asterisk along the branches represents significant statistical support for the clade subtending that branch ($p \geq 0.9$). Recent clusters potentially representing epidemiologically linked cases are indicated by letters. The tree with indicating GenBank accession numbers for tip labels is given in Supplementary Figure S5.

Figure 6. Demographic history of different HCV subtypes. The demographic history of each subtype was inferred by Bayesian skyline plots. The y-axis reports virus effective population size (*Ne*), a measure of genetic diversity representing the number of genomes effectively contributing to new infections, while the x-axis is time in calendar years. Black lines are median estimates; purple lines are 95% highest posterior density intervals. (Panel a) HCV 1b, (Panel b) HCV-3a, (Panel C) HCV-3b, and (Panel d) HCV-6a.