沖縄における C型肝炎ウイルス遺伝子型1aの分子系統解析

Phylogenetic and phylodynamic analysis of hepatitis C virus subtype 1a in Okinawa, Japan

Kunikazu Hoshino[1,2], Masaya Sugiyama[1]*, Tomoko Date[1], Shuichi Maruwaka[2], Shingo Arakaki[2], Daisuke Shibata[3], Tatsuji Maeshiro[2], Akira Hokama[2], Hiroshi Sakugawa[3], Tatsuya Kanto[4], Jiro Fujita[2], Masashi Mizokami[1]

[1]Genome Medical Science Project, National Center for Global Health and Medicine, Chiba, Japan

[2]Department of Infectious, Respiratory, and Digestive Medicine, University of the Ryukyus, Okinawa, Japan

[3]Heart Life Hospital, Digestive Division, Okinawa, Japan

[4]Department of Liver Diseases, National Center for Global Health and Medicine, Chiba, Japan

Running Head: Phylogenetic analysis of HCV-1a in Okinawa

*Address correspondence to Masaya Sugiyama

Genome Medical Sciences Project,

National Center for Global Health and Medicine, Ichikawa, Japan

1-7-1, Kohnodai, Ichikawa 272-8516, Japan.

E-mail: m.sugiyama@hospk.ncgm.go.jp

Tel:+81-(0)47-372-3501

Fax:+81-(0)47-375-4766

Abbreviations

HCV, hepatitis C virus; IDU, injecting drug use; PWID, people who inject drugs; TMRCA, time to the most recent common ancestor; MCMC, Markov chain Monte Carlo; BSP, Bayesian skyline plot; HPD, highest posterior density; NJ, neighbor-joining; MCC, maximum clade credibility; MSM, men who have sex with men; ESS, effective sample size

Abstract

Okinawa Island, located in Southern Japan, has a higher prevalence rate of hepatitis C virus subtype 1a (HCV-1a) infection than that in mainland Japan. Okinawa has a history of US military occupation after World War II. To elucidate the transmission history of HCV-1a in Okinawa, 26 whole-genome sequences were obtained from 29 patients during 2011 to 2016. Phylogenetic trees were reconstructed to identify the origin and characteristics of HCV-1a in Okinawa with epidemiological information. A phylogenetic tree based on whole-genome sequencing revealed that all of the samples were located below the United States (US) branches. Additionally, we identified one cluster comprised of 17 strains (Okinawa, n = 16; US, n = 1). The majority of the patients in this cluster were people who inject drugs (PWID), indicating the presence of a PWID cluster. Subsequently, Bayesian analyses were employed to reveal viral population dynamics. Intriguingly, a phylodynamic analysis uncovered a substantial increase in effective population size of HCV-1a from 1965 to 1980 and a slight increase in mid-2000, which were associated with an increase in illicit drug use in Okinawa. The estimated divergence time of the PWID cluster was 1967.6 (1964.2 to 1971.1). These findings suggest that HCV-1a was introduced into Okinawa from the US in the late 1960s, coincident with the Vietnam war. Subsequently, HCV-1a might have spread among the Japanese population with the spread of

injecting-drug use. Our study provides an understanding of HCV transmission dynamics in Okinawa, as well as the key role of PWID in HCV transmission.

Key Words

HCV epidemic history, HCV subytypes, Japan, Okinawa Island, phylogenetics, PWID

Introduction

Hepatitis C virus (HCV) is a widespread human pathogen and poses a major public health concern, worldwide. More than 185 million people are estimated to be seropositive for HCV worldwide (1, 2). Chronic HCV infection leads to liver fibrosis, cirrhosis, and hepatocellular carcinoma.

HCV is a positive-sense, single-stranded RNA virus that exhibits a wide range of genetic diversity. According to phylogenetic analyses, HCV is classified into seven genotypes that are further sub-divided into 67 subtypes (3). Of the HCV genotypes, HCV genotype 1 (HCV-1) infection is the most prevalent genotype worldwide, and is estimated to have infected approximately 83.4 million people (46.2% of the worldwide HCV-positive population) (2). The majority of HCV-1 subtypes consist of 1a (31%) and 1b (68%) subtypes. The distribution of HCV-1b is predominant worldwide except for North America. HCV-1a is prevalent in North America, parts of South America, the United Kingdom, Scandinavia, and Australia (2).

Based on genetic diversity, phylogenetic analyses have provided demographic and geographic estimations for the transmission history of HCV infections (4-8). HCV-1b is considered to have spread mainly via invasive medical procedures during and after World War II (9). However, the main transmission route of HCV-1a is controversial. Magiorkinis et al. reported that HCV-1a spread globally in the 1960s via high-risk behaviors such as injecting drug use (IDU), tattooing, and unprotected sex (10). However, Joy et al. estimated that the epidemic of subtype 1a in North America occurred approximately in the 1950s, suggesting that the main cause was medical procedures (11).

HCV-1a is relatively rare in Asian countries (2, 12) and only 3 whole-genome sequences of HCV-1a have been reported from the region to date. In the main islands of Japan, the prevalence

of HCV-1b is the highest, while that of HCV-1a is very low (approximately 1-2% in HCV-1 infection) (13, 14). In contrast, Okinawa Island, located in south of the main islands of Japan, has a higher prevalence rate of HCV-1a infection, which accounts for approximately 6%-24% of HCV infection among voluntary blood donors (15, 16). Okinawa has a history of the United States (US) military occupation after World War II and the spread of IDU during and after the Vietnam war. However, the origin and transmission history of HCV-1a in Okinawa remains unclear, since only limited whole-genome sequencing data regarding HCV-1a in Japan including Okinawa is available (16).

The purpose of the present study was to elucidate the origin and the transmission history of HCV-1a in Okinawa. While most previous studies used partial genome sequencing, we analyzed viral whole-genome sequencing data in combination with historical and epidemiological information to reconstruct a robust epidemic history of HCV-1a in Okinawa. We also analyzed demographic changes in individuals infected with HCV-1a in Okinawa from the perspective of people who inject drugs (PWID). Our data might be helpful for the understanding of HCV dissemination in Okinawa and provide a basis for public health strategies aimed at the control and prevention of HCV infection.

Materials and Methods

Study Population

Twenty-nine patients with serologically confirmed HCV-1a infection were enrolled in this study. All the patients were born and raised in Okinawa Island and attended the hepatology outpatient clinic of the University of the Ryukyus Hospital or its affiliated hospitals from 2011 to 2016. The study was performed in accordance with the ethical principles described in the then applicable

version of the Declaration of Helsinki (6th revision, 2008). Anonymized samples and data were offered by the University of the Ryukyus, and this study was approved by the Ethics Committee of the National Center for Global Health and Medicine. For each patient, clinical data including age, gender, and presumed routes of infection were collected.

RNA isolation, reverse transcription, PCR amplification, and sequencing

Viral RNA was extracted from 200 µl serum using the QIAamp® MinElute virus spin kit (Qiagen, Chatsworth, CA). cDNA was generated using superscript® reverse transcriptase and random primers (Invitrogen, Carlsbad, CA). The whole HCV genome, spanning the region from the 5′ untranslated region to the NS5B region (8840 nucleotides (nt); nt285-9124, according to the nucleotide numbering in the H77 reference sequence), was amplified by PCR using TaKaRa Ex Taq HS (TaKaRa, Tokyo, Japan) with a set of primer pairs (Table S1 in the supplemental material). When the amount of PCR product was insufficient, additional nested-PCR was performed using the same primer-pair. PCR conditions were 95°C for 10 min followed by 40 cycles of denaturation at 94°C for 30 s, annealing at 60°C for 30 s, extension at 72°C for 45 s with a final extension step at 72°C for 7 min. The PCR products were analyzed by agarose gel electrophoresis and were sequenced using the Sanger method with the 3500xL Genetic Analyzer (Applied Biosystems, Foster City, CA). Multiple sequence reads were assembled into a whole genome sequence using SeqMan Pro v11 (DNAStar, Madison, WI).

Preparation of data sets for phylogenetic analysis

Whole-genome sequences were successfully amplified in 26 (89.7%) out of 29 serum samples. Three samples were ruled out because of PCR amplification failure.

For the data set of reference sequences, all available whole-genome sequences of subtype 1a were collected in October 2016 from the Los Alamos National Laboratory HCV database (17).

The sequences obtained were aligned with the reference sequences using MAFFT (version 7.294b) (18) and manually inspected using MEGA (version 7.0.16) (19). Subsequently, sequences with known sampling location and date were retained, in addition to maintaining the variety of sampling locations and dates.

Two data sets of global reference sequences were prepared for whole-genome analysis. The first data set included 100 sequences with known sampling location and year, which was used for substitution rate estimation. The second data set consisted of 242 sequences with known sampling location, which comprised 6 countries: Brazil (n = 3), Switzerland (n = 23), China (n = 1), Germany (n = 13), Japan (n = 1), the US (n = 201). The second data set was used for phylogenetic analysis.

For the phylogenetic analysis of a partial HCV region, all the available reference sequences of the HCV core (573 nt; nt432-914) and NS5B (1523 nt; nt7602-9124) regions were collected for comparison with the sequences obtained from the patients from Okinawa. For the core region, 286 reference sequences were retained with information of genotype and the sampling location, which comprised 18 countries: Brazil (n = 3), Switzerland (n = 22), China (n = 1), Germany(n = 11), Denmark (n = 1), Dominican Republic (n = 1), Spain (n = 1), France (n = 1), the United Kingdom (n = 12), Iran (n = 1), Italy (n = 2), Japan (n = 3), Pakistan (n = 2), Saudi Arabia (SA) (n = 11), Sweden (n = 1), Thailand (n = 5), the US (n = 200), and Venezuela (VE) (n = 8). For the NS5B region, 377 reference sequences were obtained from a total of 8 countries: Australia (n = 21), Brazil (n = 1), Switzerland (n = 39), China (n = 1), Denmark (n = 11), Japan (n = 1), New Zealand (n = 48), and the US (n = 255).

Substitution rate

In order to estimate the substitution rate for the whole-genome region of HCV-1a, two external

rate calibration approaches were used: i) root-to-tip regression analysis and ii) Bayesian inference using a molecular clock.

For the root-to-tip regression analysis, we used 26 whole-genome sequences from Okinawa and the first data set described above. The most appropriate nucleotide substitution model for our study was estimated to be the general time-reversible model with invariable sites along with gamma-distributed among site rate variation (GTR+Γ+I) according to J-ModelTest2 (version 2.1.10) (20). Maximum likelihood (ML) trees were reconstructed using RAxML (Version 8.2.9) with the GTR+Γ+I model and 1000 bootstrap replicates (21). A linear regression analysis of root-to-tip genetic distances against the sampling dates was performed using TempEst (version 1.5) (22). The root was determined by maximizing the coefficient of the determinant, R^2.

To estimate the substitution rate in the Bayesian inference method, 26 whole-genome sequences from Okinawa and the first data set were used. Based on previous studies (23-25), we estimated the substitution rate under an uncorrelated relaxed lognormal molecular clock model with constant, exponential, and Bayesian Skyline Plot (BSP) coalescent models using BEAST (version 1.8.3) (26-29). The GTR+Γ+I model was used for the nucleotide substitution model. The nucleotide sites were partitioned into first and second codon positions and third codon positions. These models were used for all subsequent analyses. Each Markov chain Monte Carlo (MCMC) run contained 200 million states, sampled once every 10,000 states. MCMC convergence and effective sample size (ESS) were checked using Tracer (version 1.6) (28). ESS values greater than 200 were accepted after discarding the burn-in. Log marginal likelihood was calculated using Tracer (version 1.6) and the best fitting-model was selected on the basis of log Bayes Factor (30).

Phylogenetic analyses

168 To evaluate the appropriate model for phylogenetic analysis, phylogenetic trees were constructed
169 by the neighbor-joining (NJ) method as implemented in MEGA (version 7.0.16). NJ
170 bootstrapping was performed with 1000 replicates. All NJ phylogenetic trees were rooted with
171 HCV-1b sequences (AB016785, AJ132996, DQ071885, EF407479, and U01214) as the
172 outgroup using FigTree (version 1.4.2). Three types of trees for the whole-genome, core, and
173 NS5B regions were constructed by the above methods.

174 **Phylodynamic analyses and inference of divergence date**

175 The epidemic history of HCV-1a in Okinawa was estimated in Bayesian inference using the
176 MCMC algorithm as implemented in BEAST (version 1.8.3). Phylodynamic analyses were
177 carried out under a BSP coalescent model and an uncorrelated relaxed lognormal molecular
178 clock model. Each MCMC run contained 100 million states, sampled once every 10,000 states.
179 In order to disclose the divergence time of a PWID cluster in Okinawa (described later), we
180 estimated the time to the most recent common ancestor (TMRCA) using 26 sequences from our
181 study and the first data set. Maximum clade credibility (MCC) trees were constructed under a
182 constant coalescent model and an uncorrelated relaxed lognormal molecular clock model by
183 BEAST (version 1.8.3). The MCMC was run for 200 million states and sampled every 10,000
184 states.

185 **Historical data of crime related to illicit drug use in Okinawa**

186 The number of arrests and persons arrested for narcotics and stimulants use from 1961 to 2014 in
187 Okinawa was obtained from statistical data published by the Ryukyu Police department and the
188 Okinawa Prefectural Police department.

189 **Nucleotide sequence accession numbers**

190 All new sequences obtained in this study were submitted to the DNA data bank of Japan (DDBJ)

and were assigned accession numbers LC209863 to LC209888. The HCV reference sequences used in this study are shown in Table S2-S5 in the supplemental material.

Results

Determination of HCV full-genome sequences

We successfully sequenced the whole-genome of HCV-1a in 26 (89.7%) out of 29 samples. Only the HCV core and NS5B gene sequences were determined in three samples because of low viral load or coinfection. Epidemiological characteristics of the individuals are shown in Table 1. Approximately 50% of the presumed infection route involved needles, such as tattooing or IDU.

Substitution rate estimation

The substitution rate of the 26 full-genome samples was calculated using root-to-tip regression analysis by comparison with 100 reference sequences. The estimated substitution rate was 1.34×10^{-3} substitution/site/year. The TMRCA provides an estimate of when virus genes emerged in a given host, which indicates the time of interspecies transmission. The TMRCA of the samples was approximately 1932.28 (Fig. 1). However, this strict clock model is not an ideal model because it assumes that all branches evolve at exactly the same rate. In order to obtain a more accurate substitution rate, we also estimated the substitution rate using the Bayesian inference method under several conditions (Table S6). Considering the Bayes Factors, an uncorrelated relaxed lognormal clock model and a BSP coalescent model were selected. The estimated substitution rate was 1.13×10^{-3} substitution per site per year (95% highest posterior density (HPD) intervals, $9.68 \times 10^{-4} – 1.29 \times 10^{-3}$). We used this rate for subsequent analysis.

Phylogenetic analysis

Figure 2 shows the NJ phylogenetic tree estimated from the whole-genome alignments using the

26 sequences from Okinawa and the second reference data set. Interestingly, we identified a cluster that was comprised of 17 sequences, the root of which was supported by a bootstrap value > 0.9. This cluster was located below the US branch and contained one US strain (EU256015), suggesting a possible transmission route from the US. The majority of the patients in this cluster were PWID (n = 10). This PWID cluster consisted of distinct strains from Okinawa, suggesting that HCV genotype 1a transmission had occurred between Japanese people via routes that included IDU. The other strains (n = 10) were randomly dispersed in the cluster of HCV-1a, suggesting different routes of transmission: infected blood (n = 3); PWID (n = 1); tattoo (n = 1); sexual contact with foreigners (n = 1); and unknown (n = 3). When considered together, all strains obtained from Okinawa were located near or below the US strains.

Additionally, NJ phylogenetic trees were constructed for the core and NS5B regions. Sequences from Okinawa were relatively dispersed and formed small clusters in both phylogenetic trees. The phylogenetic tree based on the core region showed that 25 of 26 isolates appeared to be introduced from the US (Fig. 3A); however, the origin of one sequence could not be confirmed. In the phylogenetic tree of the NS5B region, 18 isolates from our study were located below US branches (Fig. 3B). Five sequences clustered with reference strains, which were distant from any sequence from Okinawa in the whole-genome phylogeny. The origin of another three sequences could not be determined.

Phylodynamic analyses and inference of divergence date

Figure 4A shows the BSP estimated from all of the whole-genome sequences of the HCV genotype 1a in Okinawa. A BSP is a non-parametric analysis method for estimating effective population size over time. Figure 4A indicates that gradual growth and subsequent expansion of the HCV-1a population occurred from the mid-1960s to the early 1980s. Following a plateau

phase between the mid-1980s and the early 2000s, a slight increase was identified in the mid-2000s. To date, one other plateau phase has occurred since. These effective population size data are shown above a graph that demonstrates the number of arrests and persons arrested for narcotics and stimulants use from 1961 to 2014 in Okinawa (Fig. 4B). An explosive increase in narcotics-related crimes in the early 1970s is associated with the growth of the effective population size of HCV-1a in the 1970s. Likewise, a gradual increase in stimulants-related crimes in the 2000s is consistent with the growth of the effective population size in the mid-2000s.

To estimate the divergence time of the PWID cluster in Okinawa, a phylogenetic analysis of the MCC molecular clock was performed using the HCV full-genome sequences. The most recent common ancestor of the PWID cluster was found in the year 1967.60 (95% HPD intervals, 1964.20 to 1971.05) (Fig. 5).

Discussion

While HCV-1 is the most prevalent HCV genotype worldwide, HCV-1a is relatively rare in Asian countries including Japan. In contrast, Okinawa Island, which is located in Southern Japan, has a higher prevalence rate of HCV-1a infection than that for the rest of Japan. However, little is known about the transmission history of HCV-1a strains in Okinawa. In order to investigate a history of HCV-1a transmission in Okinawa, we performed fine-scale phylogenetic analyses based on informative whole-genome sequencing data with historical and epidemiological information.

Among Asian countries, Thailand, the Philippines, and Vietnam have a relatively high prevalence of HCV-1a. Previous studies speculated that HCV-1a prevalence in those countries might be

associated with the deployment of US military forces in the past (31, 32). Additionally, one HCV-1a strain that clustered with US strains was detected from the Bonin Islands (the Ogasawara Islands, Japan), which were also occupied by US military forces until 1968 (16). Similar to the results of these studies, our results showed that most of the HCV-1a strains circulating in Okinawa were likely to have originated in the US. A phylodynamic study revealed a substantial increase in effective HCV-1a population size from 1965 to 1980. In addition, the estimated date of the divergence time for the PWID cluster was approximately 1968 (1964 to 1971), and this estimated time was similar under different coalescent models. The estimated starting point of the increase, the mid-1960s to the early 1970s, coincides with the latter part of the Vietnam war. Okinawa has a history of US military occupation after World War II. Even after Okinawa was returned to Japan in 1972, Okinawa has had a strong socioeconomic relationship with the US. According to a previous study, IDU spread among Japanese people in Okinawa from the late 1960s to the mid-1970s (33). Around the same time, a considerable prevalence of heroin use was reported among US military servicemen (34, 35), and the same trend was also reported in Okinawa (36). These historical backgrounds and the results of our phylogenetic analyses suggest that contact with people from the US played an important role in the introduction of HCV-1a into Okinawa.

IDU is one of the highest risk factors among transmission routes for HCV infection, with an odds ratio of 49.6 (37). In addition, PWID are considered as a major origin of the persistent HCV epidemic (38). In this study, we identified a PWID cluster that included 16 strains from Okinawa. We also demonstrated that the rapid growth in effective population size of HCV-1a was consistent with the explosive increase in narcotics-related crimes in the 1970s. These findings suggest that some of the HCV-1a spread among the Japanese population in Okinawa during and

283 after the Vietnam war. Furthermore, we speculate that IDU might have been the driving-force
284 behind the spread of HCV-1a in Okinawa.
285 Although the number of narcotics-related crimes decreased until the mid-1970s, the number of
286 stimulants-related crimes in Okinawa gradually increased from the late 1970s and increased
287 further in the year 2000 (Fig. 4B). These historical records support our phylodynamic results that
288 the effective population size of HCV-1a infection increased in Okinawa in the mid-2000s. A
289 similar increase was reported in an analysis of HCV-1a samples from the North American
290 population (10). Almost all developed countries had a problem related to an increase in young
291 drug abusers and HCV carriers around the year 2000 (39). The PWID population in Okinawa
292 also includes such young drug users, and this active population could have been a cause of HCV
293 spread in the mid-2000s. Effective strategies aimed at this population are needed to reduce the
294 future burden of HCV-related liver disease.
295 Previous studies reported that HCV-2 or 3, but not HCV-1, was prevalent among PWID in the
296 main islands of Japan (40-44). These data suggest that HCV-2/3 infection occurred by another
297 distribution channel of illegal drugs in the main islands of Japan. We have revealed a large-scale
298 ongoing HCV-1a transmission network among the Japanese high-risk population, which has not
299 been reported previously. However, the extent of HCV-1a transmission among PWID in Japan
300 and its surrounding countries is unclear. It is possible that there are other unrecognized
301 transmission networks and/or risk factors. Further study is needed for better understanding of
302 HCV-1a spread in Okinawa.
303 Unfortunately, there were some probable transmission routes in the enrolled population that we
304 did not analyze. Although men who have sex with men (MSM), particularly HIV-positive
305 individuals, have a high risk of HCV infection (45), they were not included in our patients based

306 on our interview. Therefore, we have not evaluated the impact of MSM behavior on the spread of
307 HCV infection. Another limitation of this study is the small sample size used to perform the
308 phylogenetic analyses for an evolutionary investigation.
309 However, we believe that analysis of 26 full-genome sequences of HCV-1a in Japan enabled us
310 to investigate the phylogenetic relationships with relative accuracy, and this is the first such
311 report from Japan. Our present data could provide useful information on the transmission route
312 and could support planning regarding the control and prevention of HCV infection. Notably, we
313 revealed for the first time a large-scale HCV transmission network among the Japanese high-risk
314 population. A future comprehensive investigation would be important to reveal the presence and
315 characteristics of patients with ongoing HCV infection.
316 While most previous studies used the HCV core and NS5B regions for phylogenetic analyses, we
317 instead performed HCV whole-genome sequencing. Although some authors have proposed the
318 E1, E2, NS2, NS4B (11), and E2P7NS2 (10) regions as being the more informative regions,
319 which viral genome regions provide the best phylogenetic trees is controversial. The topology of
320 the PWID cluster of our study varied significantly in phylogenetic trees based on the different
321 genome regions, indicating that analyses based on a partial genome region could distort the
322 structure of the HCV transmission network. In a comparison of topologies derived from different
323 combinations of HIV viral genes, Yebra et al. showed that the accuracy of HIV phylogenetic
324 trees was significantly associated with the length of the viral sequences (46). A finer and more
325 accurate phylogenetic reconstruction could be achieved through more informative molecular data.
326 Alongside epidemiological information and social network data (47), whole-genome viral data
327 would further contribute to the determination of transmission pathways and to the development
328 of effective strategies for the prevention and control of HCV infection.

Although the MSM population was not included in our study, it has been reported that the number of HIV cases among MSM has increased in Okinawa (48). The HCV transmission network among MSM is international (45, 49, 50) and MSM tourists visit Okinawa from other areas in Japan or overseas (51). These data suggest that HCV-1a strains could possibly spread out from the HIV-positive MSM population in Okinawa into the MSM population in the main islands of Japan or in other Asian countries. Further research is needed to evaluate HCV transmission among this high-risk population from an international public health perspective.

In conclusion, we investigated the dynamics of HCV-1a transmission in Okinawa by combining historical and epidemiological information with viral whole-genome data. The first HCV-1a infection in Okinawa could have occurred via contact with people from the US during the time of the Vietnam War through various routes. Subsequently, the HCV infection may have spread among Japanese individuals with high-risk behaviors. Notably, our study revealed for the first time a large-scale HCV transmission network in Japan, for which control and prevention strategies should be developed from the viewpoint of public health. Our study sheds light on the dynamics of HCV transmission and provides a basis for its effective control.

Acknowledgment and disclosures

We have no conflicts of interest.

References

1. Mohd Hanafiah K, 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(4):1333-42.

2. 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.

3. 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.

4. Mizokami M, Gojobori T, Lau JY. Molecular evolutionary virology: its application to hepatitis C virus. Gastroenterology. 1994;107(4):1181-2.

5. Tanaka Y, Hanada K, Mizokami M, Yeo AE, Shih JW, Gojobori T, et al. A comparison of the molecular clock of hepatitis C virus in the United States and Japan predicts that hepatocellular carcinoma incidence in the United States will increase over the next two decades. Proc Natl Acad Sci U S A. 2002;99(24):15584-9.

6. Tanaka Y, Hanada K, Orito E, Akahane Y, Chayama K, Yoshizawa H, et al. Molecular evolutionary analyses implicate injection treatment for schistosomiasis in the initial hepatitis C epidemics in Japan. J Hepatol. 2005;42(1):47-53.

7. 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.

8. Fu Y, Qin W, Cao H, Xu R, Tan Y, Lu T, et al. 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(1):e28006.

9. Simmonds P. The origin of hepatitis C virus. Curr Top Microbiol Immunol. 2013;369:1-15.

10. Magiorkinis G, Magiorkinis E, Paraskevis D, Ho SY, Shapiro B, Pybus OG, et al. The global spread of hepatitis C virus 1a and 1b: a phylodynamic and phylogeographic analysis.

375 PLoS Med. 2009;6(12):e1000198.
376 11. Joy JB, McCloskey RM, Nguyen T, Liang RH, Khudyakov Y, Olmstead A, et al. The
377 spread of hepatitis C virus genotype 1a in North America: a retrospective phylogenetic study.
378 Lancet Infect Dis. 2016;16(6):698-702.
379 12. Wei L, Rao H-Y, Wang Y, Yang M, Gao Y-H. Molecular Epidemiology of HCV in Asia.
380 Current Hepatitis Reports. 2013;12(3):133-42.
381 13. Hayashi K, Katano Y, Kuzuya T, Tachi Y, Honda T, Ishigami M, et al. Prevalence of
382 hepatitis C virus genotype 1a in Japan and correlation of mutations in the NS5A region and
383 single-nucleotide polymorphism of interleukin-28B with the response to combination therapy
384 with pegylated-interferon-alpha 2b and ribavirin. J Med Virol. 2012;84(3):438-44.
385 14. Wu S, Kanda T, Nakamoto S, Jiang X, Miyamura T, Nakatani SM, et al. Prevalence of
386 hepatitis C virus subgenotypes 1a and 1b in Japanese patients: ultra-deep sequencing analysis of
387 HCV NS5B genotype-specific region. PLoS One. 2013;8(9):e73615.
388 15. Sakugawa H, Nakasone H, Kinjo F, Saito A, Keida Y, Kikuchi K, et al. Clinical
389 features of patients with chronic liver disease associated with hepatitis C virus genotype 1a/I in
390 Okinawa, Japan. J Gastroenterol Hepatol. 1997;12(2):176-81.
391 16. Okamoto H, Kanai N, Mishiro S. Full-length nucleotide sequence of a Japanese
392 hepatitis C virus isolate (HC-J1) with high homology to USA isolates. Nucleic Acids Res.
393 1992;20(23):6410.
394 17. Kuiken C, Yusim K, Boykin L, Richardson R. The Los Alamos hepatitis C sequence
395 database. Bioinformatics. 2005;21(3):379-84.
396 18. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7:
397 improvements in performance and usability. Mol Biol Evol. 2013;30(4):772-80.

398 19. Kumar S, Stecher G, Tamura K. MEGA7: Molecular Evolutionary Genetics Analysis
399 Version 7.0 for Bigger Datasets. Mol Biol Evol. 2016;33(7):1870-4.
400 20. Darriba D, Taboada GL, Doallo R, Posada D. jModelTest 2: more models, new
401 heuristics and parallel computing. Nat Methods. 2012;9(8):772.
402 21. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of
403 large phylogenies. Bioinformatics. 2014;30(9):1312-3.
404 22. Rambaut A, Lam TT, Max Carvalho L, Pybus OG. Exploring the temporal structure of
405 heterochronous sequences using TempEst (formerly Path-O-Gen). Virus Evol.
406 2016;2(1):vew007.
407 23. Raghwani J, Thomas XV, Koekkoek SM, Schinkel J, Molenkamp R, van de Laar TJ, et
408 al. Origin and evolution of the unique hepatitis C virus circulating recombinant form 2k/1b. J
409 Virol. 2012;86(4):2212-20.
410 24. Gray RR, Parker J, Lemey P, Salemi M, Katzourakis A, Pybus OG. The mode and
411 tempo of hepatitis C virus evolution within and among hosts. BMC Evol Biol. 2011;11:131.
412 25. Yuan M, Lu T, Li C, Lu L. The evolutionary rates of HCV estimated with subtype 1a
413 and 1b sequences over the ORF length and in different genomic regions. PLoS One.
414 2013;8(6):e64698.
415 26. Drummond AJ, Rambaut A, Shapiro B, Pybus OG. Bayesian coalescent inference of
416 past population dynamics from molecular sequences. Mol Biol Evol. 2005;22(5):1185-92.
417 27. Drummond AJ, Ho SY, Phillips MJ, Rambaut A. Relaxed phylogenetics and dating
418 with confidence. PLoS Biol. 2006;4(5):e88.
419 28. Drummond AJ, Rambaut A. BEAST: Bayesian evolutionary analysis by sampling trees.
420 BMC Evol Biol. 2007;7:214.

421 29. Drummond AJ, Suchard MA, Xie D, Rambaut A. Bayesian phylogenetics with
422 BEAUti and the BEAST 1.7. Mol Biol Evol. 2012;29(8):1969-73.
423 30. Suchard MA, Weiss RE, Sinsheimer JS. Bayesian selection of continuous-time Markov
424 chain evolutionary models. Mol Biol Evol. 2001;18(6):1001-13.
425 31. Li C, Yuan M, Lu L, Lu T, Xia W, Pham VH, et al. The genetic diversity and
426 evolutionary history of hepatitis C virus in Vietnam. Virology. 2014;468-470:197-206.
427 32. Akkarathamrongsin S, Hacharoen P, Tangkijvanich P, Theamboonlers A, Tanaka Y,
428 Mizokami M, et al. Molecular epidemiology and genetic history of hepatitis C virus subtype 3a
429 infection in Thailand. Intervirology. 2013;56(5):284-94.
430 33. Taira H. Studies on Narcotic Crimes and Narcotic Addicts in Okinawa. Kurume Med J.
431 1982;29(3):141-55.
432 34. Baker SL, Jr. Drug abuse in the United States Army. Bull N Y Acad Med.
433 1971;47(6):541-9.
434 35. Robins LN, Slobodyan S. Post-Vietnam heroin use and injection by returning US
435 veterans: clues to preventing injection today. Addiction. 2003;98(8):1053-60.
436 36. Baker SL, Jr. Present status of the drug abuse counteroffensive in the Armed Forces.
437 Bull N Y Acad Med. 1972;48(5):719-32.
438 37. Murphy EL, Bryzman SM, Glynn SA, Ameti DI, Thomson RA, Williams AE, et al.
439 Risk factors for hepatitis C virus infection in United States blood donors. NHLBI Retrovirus
440 Epidemiology Donor Study (REDS). Hepatology. 2000;31(3):756-62.
441 38. Edlin BR, Carden MR. Injection drug users: the overlooked core of the hepatitis C
442 epidemic. Clin Infect Dis. 2006;42(5):673-6.
443 39. Page K, Morris MD, Hahn JA, Maher L, Prins M. Injection drug use and hepatitis C

444 virus infection in young adult injectors: using evidence to inform comprehensive prevention.
445 Clin Infect Dis. 2013;57 Suppl 2:S32-8.
446 40. Chung H, Ueda T, Kudo M. Changing trends in hepatitis C infection over the past 50
447 years in Japan. Intervirology. 2010;53(1):39-43.
448 41. Morikawa H, Kozuka R, Fujii H, Iwai S, Enomoto M, Tamori A, et al. Combination
449 therapy of natural human interferon-beta and ribavirin for chronic hepatitis C patients with
450 injection drug use. Hepatol Res. 2013;43(10):1013-9.
451 42. Satoh Y, Hino K, Kato T, Mizokami M, Yamashita S, Nakamura H, et al. Molecular
452 epidemiologic analysis of hepatitis C virus infection in injecting drug users with acute hepatitis C
453 in Japan. J Gastroenterol Hepatol. 2004;19(11):1305-11.
454 43. Kato H, Maeno Y, Seko-Nakamura Y, Monma-Ohtaki J, Sugiura S, Takahashi K, et al.
455 Identification and phylogenetic analysis of hepatitis C virus in forensic blood samples obtained
456 from injecting drug users. Forensic Sci Int. 2007;168(1):27-33.
457 44. Robaeys G, Bielen R, Azar DG, Razavi H, Nevens F. Global genotype distribution of
458 hepatitis C viral infection among people who inject drugs. J Hepatol. 2016;65(6):1094-103.
459 45. Midgard H, Weir A, Palmateer N, Lo Re V, 3rd, Pineda JA, Macias J, et al. HCV
460 epidemiology in high-risk groups and the risk of reinfection. J Hepatol. 2016;65(1
461 Suppl):S33-45.
462 46. Yebra G, Hodcroft EB, Ragonnet-Cronin ML, Pillay D, Brown AJ, Consortium PH, et
463 al. Using nearly full-genome HIV sequence data improves phylogeny reconstruction in a
464 simulated epidemic. Sci Rep. 2016;6:39489.
465 47. Vasylyeva TI, Friedman SR, Paraskevis D, Magiorkinis G. Integrating molecular
466 epidemiology and social network analysis to study infectious diseases: Towards a

socio-molecular era for public health. Infect Genet Evol. 2016;46:248-55.

48. Lim SH, Chan R. HIV infection among men who have sex with men in East and South-East Asia - time for action. Sex Health. 2011;8(1):5-8.

49. van de Laar T, Pybus O, Bruisten S, Brown D, Nelson M, Bhagani S, et al. Evidence of a Large, International Network of HCV Transmission in HIV-Positive Men Who Have Sex With Men. Gastroenterology. 2009;136(5):1609-17.

50. Takebe Y, Naito Y, Raghwani J, Fearnhill E, Sano T, Kusagawa S, et al. Intercontinental dispersal of HIV-1 subtype B associated with transmission among men who have sex with men in Japan. J Virol. 2014;88(17):9864-76.

51. Sherriff N, Koerner J, Kaneko N, Shiono S, Takaku M, Boseley R, et al. Everywhere in Japan: an international approach to working with commercial gay businesses in HIV prevention. Health Promot Int. 2015.

Supplementary material

Table S1: Primers for PCR of HCV

Table S2: HCV whole-genome sequences used in this study (The first data set, n = 100)

Table S3: HCV whole-genome sequences used in this study (The second data set, n = 242)

Table S4: HCV core region sequences used in this study (n = 286)

Table S5: HCV NS5B region sequences used in this study (n = 377)

Table S6: Estimated substitution rate results for 26 whole-genome strains from Okinawa and 100 reference sequences

489 **TABLE 1** Epidemiological characteristics of the patients enrolled in this study.

No	Age (year)	Gender	Sampling year	Presumed infection route
1	66	M	2016	Unknown
2	62	M	2015	Tattoo
3	61	F	2014	Sexual contact with foreigners
4	57	F	2015	Sexual contact with foreigners
5	59	F	2015	Unknown
6	64	M	2015	Unknown
7	29	M	2015	Blood products
8	72	F	2015	Transfusion
9	54	M	2016	Sexual contact with foreigners /Tattoo
10	60	M	2015	Sexual contact with foreigners /Contact with female sex workers
11	39	M	2016	PWID[a]
12	40	F	2016	PWID
13	64	M	2015	PWID
14	58	F	2014	Blood products
15	34	M	2011	PWID
16	29	M	2015	PWID /Tattoo
17	58	F	2015	Unknown
18[b]	65	F	2016	Unknown
19	61	M	2016	PWID

No	Age (year)	Gender	Sampling year	Presumed infection route
20	25	F	2015	Needlestick injury
21	61	F	2016	Sexual contact with foreigners
22	65	F	2016	PWID
23	61	F	2016	Unknown
24	48	F	2014	PWID
25	62	F	2016	PWID
26 [b]	35	M	2007	PWID
27	56	M	2015	PWID
28	44	M	2016	PWID
29 [b]	50	M	2014	Unknown

490 [a] PWID, people who inject drugs

491 [b] Not available due to PCR amplification failure

492

Figure legends

Figure 1

Root-to-tip regression analysis using maximum likelihood trees for the whole-genome region. Twenty-six sequences sampled in Okinawa were analyzed with 100 global reference sequences using TempEst (version 1.5). The root was determined by maximizing the coefficient of the determinant, R^2. The mean evolutionary rate is 1.34×10^{-3} substitution/site/year. The estimated time to the most recent common ancestor (TMRCA) is 1932.28.

Figure 2

The neighbor-joining (NJ) phylogenetic tree based on whole-genome sequencing. Twenty-six sequences from Okinawa and 242 reference sequences were analyzed with HCV/1b strains (AB016785, AJ132996, DQ071885, EF407479, and U01214) as an outgroup using MEGA (version 7.0.16). NJ bootstrapping was performed with 1000 replicates. Tree Branches are colored according to inferred locations (blue, Okinawa; purple, the United States (US); green, other countries including Brazil, Switzerland, China, Germany, and Japan). The branch length is indicated in the units of nucleotide substitutions per site by the scale at the bottom of the tree. The outgroup is colored black. The PWID cluster, including 16 strains form Okinawa and 1 isolate from the US, are shown expanded in the box at right. The bootstrap values are shown next to the branches. Taxon labels include subtype, sampling location using two letter country codes (ISO 3166), sampling year, sequence name, and accession number. The age, gender, and presumed transmission routes are indicated next to the taxon. Abbreviations: PWID, people who inject drugs; Foreigners, sexual contact with foreigners; FSWs, female sex workers.

Figure 3

The neighbor-joining (NJ) phylogenetic trees based on the HCV core (A) and NS5B (B) region. Twenty-six sequences from Okinawa and global reference sequences were analyzed with HCV-1b strains as an outgroup using MEGA (version 7.0.16). NJ bootstrapping was performed with 1000 replicates. Tree branches are colored according to inferred locations (blue, Okinawa; purple, the United States (US); green, other countries). The branch length is indicated in the units of nucleotide substitutions per site by the scale at the bottom of the trees. Internal branches are colored gray, when their locations are unclear. The outgroup is colored black. Clades containing reference sequences from the same sampling location are collapsed. The numbers next to phylogeny tips denote the sample number in our study. The samples belonging to the PWID cluster are colored red.

Figure 4

Estimated effective population size of HCV-1a infections in Okinawa and the numbers of arrests and persons arrested for narcotics and stimulants use in Okinawa over time. (A) Bayesian Skyline Plot estimated from 26 whole-genome sequences of HCV-1a in Okinawa using BEAST (version 1.8.3). The thick black line represents the estimated effective number of infections over time. The blue area represents the 95% highest posterior density confidence intervals of this estimate. (B) The number of arrests and persons arrested for narcotics and stimulants use from 1961 to 2014 in Okinawa. This graph is based on statistical data from the Ryukyu Police department and the Okinawa Prefectural Police department.

Figure 5

539 Maximum clade credibility tree for estimating the divergence time of the PWID cluster in
540 Okinawa. Using the 26 sequences obtained in our study and 100 reference sequences, the time
541 to the most recent common ancestor (TMRCA) of the PWID cluster was estimated under a
542 constant coalescent model and an uncorrelated relaxed lognormal molecular clock. The 95%
543 highest posterior density (HPD) interval on node height estimates are shown as blue bars on
544 nodes. The most recent common ancestor of the PWID cluster is colored red and the node
545 height and 95% HPD interval (indicated inside parentheses) are shown next to the node.
546 Taxons are labeled as in Figure 2.
547