Title

Historical Trends in the Hepatitis C Virus Epidemics in North America and Australia.

Permalink

https://escholarship.org/uc/item/6161089g

Journal

The Journal of Infectious Diseases, 214(9)

Authors

Rodrigo, Chaturaka

Eltahla, Auda

Bull, Rowena

et al.

Publication Date

2016-11-01

DOI

10.1093/infdis/jiw389

Peer reviewed

Historical Trends in the Hepatitis C Virus Epidemics in North America and Australia

Chaturaka Rodrigo,[1] Auda A. Eltahla,[1] Rowena A. Bull,[1] Jason Grebely,[2] Gregory J. Dore,[2] Tanya Applegate,[2] Kimberly Page,[4] Julie Bruneau,[9] Meghan D. Morris,[5] Andrea L. Cox,[6] William Osburn,[6] Arthur Y. Kim,[7] Janke Schinkel,[10] Naglaa H. Shoukry,[9] Georg M. Lauer,[7] Lisa Maher,[2] Margaret Hellard,[3] Maria Prins,[10,11] Chris Estes,[8] Homie Razavi,[8] Andrew R. Lloyd,[1,2] and Fabio Luciani[1]; on behalf of the International Collaborative of Incident HIV and Hepatitis C in Injecting Cohorts (InC3) Study Group

[1]School of Medical Sciences, Faculty of Medicine, [2]The Kirby Institute, UNSW Australia, Sydney, New South Wales, and [3]Burnet Institute, Melbourne, Victoria, Australia; [4]University of New Mexico, Albuquerque; [5]Department of Epidemiology and Biostatistics, University of California, San Francisco; [6]Department of Medicine, Johns Hopkins Medical Institutions, Baltimore, Maryland; [7]Harvard Medical School, Boston, Massachusetts; [8]Center for Disease Analysis, Louisville, Colorado; [9]CRCHUM, Université de Montréal, Quebec, Canada; [10]Department of Internal Medicine, Division of Infectious Diseases, Tropical Medicine and AIDS, Center for Infection and Immunity Amsterdam, Academic Medical Center, Meibergdreef, and [11]GGD Public Health Service of Amsterdam, The Netherlands

Background. Bayesian evolutionary analysis (coalescent analysis) based on genetic sequences has been used to describe the origins and spread of rapidly mutating RNA viruses, such as influenza, Ebola, human immunodeficiency virus (HIV), and hepatitis C virus (HCV).

Methods. Full-length subtype 1a and 3a sequences from early HCV infections from the International Collaborative of Incident HIV and Hepatitis C in Injecting Cohorts (InC3), as well as from public databases from a time window of 1977–2012, were used in a coalescent analysis with BEAST software to estimate the origin and progression of the HCV epidemics in Australia and North America. Convergent temporal trends were sought via independent epidemiological modeling.

Results. The epidemic of subtype 3a had more recent origins (around 1950) than subtype 1a (around 1920) in both continents. In both modeling approaches and in both continents, the epidemics underwent exponential growth between 1955 and 1975, which then stabilized in the late 20th century.

Conclusions. Historical events that fuelled the emergence and spread of injecting drug use, such as the advent of intravenous medical therapies and devices, and growth in the heroin trade, as well as population mixing during armed conflicts, were likely drivers for the cross-continental spread of the HCV epidemics.

Keywords. BEAST; coalescent; epidemiology; HCV; injecting drug use; PWID.

Hepatitis C is a chronic viral infection predominantly spread via blood-to-blood contamination during injecting drug use, with outbreaks also attributed to nonsterile medical practices, unscreened blood transfusions and sexual transmission among men who have sex with men [1–5]. The hepatitis C virus (HCV) genera are divided into 7 genotypes and 67 subtypes [6]. The subtypes that have a global presence are 1a, 1b, 2a, 2b, and 3a [7]. Given the limited repertoire of modes of transmission, HCV provides an ideal opportunity to study how historical trends in public health measures, risk behaviors (eg, injecting drug use), and major socioeconomic changes such as wars and migration patterns may have shaped the global epidemic.

Viral sequences and phylogenetic analysis have been used previously to infer evolutionary dynamics of Ebola and human immunodeficiency virus (HIV), as well as HCV epidemics [1, 8]. Because these RNA viruses feature high mutation rates, the evolutionary and epidemiological time scales are similar.

These viruses are therefore termed "measurably evolving populations" [9]. In evolutionary analysis, the extent to which the sampled genealogy (ie, the ancestral lineage of the virus) represents the true genealogy of the population (and hence the accuracy of estimated parameters) depends on several factors: first, the similarity between an experimentally reconstructed and the true genealogy; second, the length of the viral sequences (phylogenetic signal); third, the model of evolution used; and finally, the phylogenetic techniques used to reconstruct the genealogy [9].

We have recently established the largest repository of early infection, full-length, HCV genome sequences available globally [10]. The contributions to this repository came from 9 prospective cohorts constituting the International Collaboration of Incident HIV and Hepatitis C in Injecting Cohorts (InC3) [11]. Phylogenetic analysis of these sequences has shown that North American and Australian sequences clustered separately from each other, suggesting that HCV largely evolves separately in microepidemics in geographically separate communities [10]. Here, we explore this theory further using Bayesian evolutionary analysis to model the spread of the subtypes 1a and 3a HCV epidemics in North America and Australia in the 20th century to understand the similarities and differences in the epidemics. This is the first description of historical trends in the HCV epidemic for subtypes 1a and 3a in Australia, and for

subtype 3a in North America using full-length genomes to maximize sensitivity in phylogenetic analyses.

METHODS

Selection of Sequences

From 368 early incident samples from InC3, 190 full-length and 23 partial-length HCV sequences were generated using a protocol of reverse-transcription and nested PCR [12]. All amplicons were subjected to next-generation sequencing (MiSeq; Illumina). The sequencing polymerase chain reaction had a mean coverage of 17 000 readings per site. The consensus sequences for each sample were generated using the bioinformatics pipeline described elsewhere [10]. The current analysis used all of the consensus sequences originating from samples collected in Australia and North America (Canada and United States) (n = 161) for subtypes 1a and 3a (4 data sets). These incident infection samples had been collected between 1998 and 2012. Public databases (Los Alamos HCV sequence database and GenBank) were also searched for full-length HCV sequences originating from North America or Australia with a known date of sampling. For subtype 3a, there were no additional full-length sequences for Australia, but an additional 6 sequences were found for North America (2 from the United States and 4 from Canada). By contrast, >700 full-length sequences were available for subtype 1a from the United States. From these additional sequences, a randomly selected subset was included to represent the time window from 1977 to 1998 to supplement the more recent InC3 sequences. The composition of the final data sets is shown in Table 1 and in the Supplementary material (S1). Each data set was aligned with MUSCLE (MUltiple Sequence Comparison by Log-Expectation) algorithm. Ethics approval for the InC3 sequencing project was granted by the UNSW Australia Human Research Ethics Committee (No. 14201), and included local ethical approval at each contributing site.

Coalescent Analysis

The analysis was completed using Bayesian evolutionary methods with BEAST software (version 1.8) [13]. The alignments were analyzed using MEGA software (version 6.0) to identify the best substitution model to explain the observed mutations [14]. The generalized time-reversible model (GTR + G + I) was selected as the best model. The most suitable clock model for the analysis was selected after running the analysis separately using strict, log-normal relaxed, and exponential relaxed models for 2 data sets of genotype 1a and 3a. The log-normal relaxed clock was selected as the best model from marginal likelihood analysis conducted by calculating posterior simulation-based analogue of Akaike information criterion (data not shown) [15, 16]. substitution rates for the full genome, previously estimated by Gray et al [17] and Choudhary et al [15] for subtypes 1a and 3a, respectively, were used as external priors in this analysis. The SRD06 model was used to estimate relative

Table 1. Selection of Sequences for the Coalescent Analysis by Continent and Subtype

Continent and Subtype	Source	Time Window, Calendar Year	Sequences, No.
North America			
1a	InC3	1998–2011	58
	Public	1977–1998	14
	Total	. . .	72
3a	InC3	2000–2011	35
	Public	1991–2000	7
	Total	. . .	42
Australia			
1a	InC3	2004–2012	30
	Public	Not available	0
	Total	. . .	30
3a	InC3	2005–2012	35
	Public	Not available	0
	Total	. . .	35

Abbreviation: InC3, International Collaborative of Incident HIV and Hepatitis C in Injecting Cohorts.

substitution rates, the α parameter, and the proportion of invariant sites for the first 2 codon positions and the third codon position separately. The Bayesian skyline plot approach was used to infer population dynamics of the epidemic [18].

The simulations were run with a Monte Carlo Markov Chain (MCMC) until the chain entered an area of high probability for each parameter. For each data set, the MCMC was run 50–500 million times for convergence. All estimates were drawn from a pool with an effective sample size >200. The MCMC chains were visualized using TRACER software (version 1.6) [19]. The tree log file was annotated using Tree Annotator software [18] and the maximum clade credibility tree was visualized with the Figtree (Andrew Rambaut, University of Edinburgh, UK, Version 1.4.2). Bayesian skyline plots were generated from the BEAST software output to visualize the changes in the effective population size since the time of the most recent common ancestor (tMRCA) to the most recent sampling time point.

Epidemiological Model

A previously developed epidemiological model was applied to estimate the time course of changes in the incidence and prevalence of HCV infection [20, 21]. The Markov model (implemented in Microsoft Excel [2013]) was used to predict the HCV-infected population and the disease progression from 1950 to 2012. The mean number of new infections per year was estimated using the known prevalence estimates in the United States [22, 23] and Australia (see below), after taking into account the trends in all-cause mortality, excess mortality among persons who inject drugs, estimated liver-related deaths, and the impact of treatment. The change in incidence was estimated by distributing the mean number of new infections between 1950 and the year of known prevalence to match the

Table 2. Summary of Coalescent Analysis for Subtypes 1a and 3a for North America and Australia

Variable	Subtype 1a		Subtype 3a	
	Australia	North America	Australia	North America
tMRCA (95% HPD), y	97.12 (50.88–149.38)	85.68 (75.11–96.32)	59.69 (43.31–77.53)	61.59 (50.65–73.51)
tMRCA (95% HPD), calendar year	1915 (1863–1961)	1925 (1914–1936)	1953 (1934–1968)	1949 (1937–1960)
Ucld mean	1.13×10^{-3}	1.07×10^{-3}	1.19×10^{-3}	1.16×10^{-3}
COV	0.586	0.123	0.073	0.111
α C(1 + 2)	0.529	0.384	0.450	0.435
α C(3)	1.521	1.282	1.358	1.423
pInV C(1 + 2)	0.754	0.694	0.732	0.738
pInV C(3)	0.120	0.075	0.029	0.048
μ C(1 + 2)	0.391	0.513	0.479	0.516
μ C(3)	2.218	1.975	2.041	1.968
MCMC length	400 million	500 million	50 million	400 million

Abbreviations: μ, relative substitution rate; C(1 + 2), relative substitution rate of first and second codon positions taken together; C(3), relative substitution of third codon position; COV, coefficient of variation; HPD, high posterior density; MCMC, Monte Carlo Markov chain; pInv, proportion of invariant sites; tMRCA, time of the most recent common ancestor; Ucld mean, estimated substitution rate.

age distribution of the known prevalence. Uncertainty and sensitivity analyses were also performed using Crystal Ball software, an Excel (2013) add-in by Oracle. Beta-PERT distributions were used to model uncertainty associated with all inputs. The model was simulated using a Monte-Carlo approach to determine the 95% uncertainty interval for prevalence [20, 21].

The epidemiological estimates of HCV infection prevalence by age and sex [22–24] for the United States were extracted from the National Health and Nutrition Examination Survey, the US Centers for Disease Control and Prevention surveillance data between 1997 and 2007, and the number of patients with HCV infection cured antiviral therapy reported by publications between 1997 and 2007 [20]. The model was also calibrated to account for US populations not captured by the National Health and Nutrition Examination Survey, such as incarcerated and homeless persons. For Canada, data reported by Public Health Agency Canada on HCV notifications between 1998 and 2011 were used to calibrate the model [25]. For Australia, data from the National Health Surveillance system between 1995 and 2013 were used [26]. These epidemiological estimates of prevalence were then compared with the changes of effective population size over time inferred by the Bayesian skyline model. The epidemiological model estimates the prevalence of HCV infection as the total infected population over time in a given region, whereas the Bayesian estimates represent the changes in effective population size, a simulated population based on the available sequences. These 2 quantities are not identical but the trends in these 2 estimates may be compared [9].

RESULTS

Coalescent Analysis

The evolutionary analysis aimed to identify the origins of the subtype 1a and 3a epidemics in both North America and Australia and the historical progression since the origin. Results are summarized in Table 2. The subtype 1a epidemic was estimated to have originated at the beginning of the 20th century, around 1915 (95% high posterior density [HPD], 1863–1961) and 1925 (95% HPD, 1914–1936) for Australia and North America, respectively. HPD interval for the 1a epidemic in Australia was wider, owing to the lack of availability of sequences dated closer to the origin of the epidemic compared with other data sets. The subtype 3a epidemic had more recent origins than the subtype 1a epidemic in both continents. The Australian 3a epidemic was estimated to have originated around 1953 (95% HPD, 1934–1968), and the North American epidemic around 1949 (95% HPD, 1937–1960). The estimated substitution rates for genotypes 1a and 3a in both continents (Australia, 1.13×10^{-3} and 1.19×10^{-3}, respectively; North America, 1.07×10^{-3} and 1.16×10^{-3}) were similar, despite the difference in genotype-specific prior rates (1.48×10^{-3} for genotype 1a and 1.07×10^{-3} for genotype 3a).

History of the Epidemics in Australia and North America

Reconstructions of the epidemics for each subtype and continent are shown on the Bayesian skyline plots in Figure 1. After the origin of the epidemic of subtype 1a in the early 20th century in North America, the effective population size remained low until 1955. Thereafter, 2 phases of rapid growth occurred between 1955 and 1975. The effective population size stabilized thereafter and declined slightly after the 1990s. The subtype 1a epidemic for Australia also showed a period of rapid growth between 1955 and 1975. However, the effective population size was in decline from the 1980s, with a faster drop observed after year 2000.

By contrast, the subtype 3a epidemics in both continents had more recent origins compared with subtype 1a. The 3a epidemic in North America showed an exponential increase in the period between 1955 and 1970. The numbers started to decline slightly after 1995. In Australia, the rapid rise in infections happened between 1960 and 1975, before stabilization through to the recent
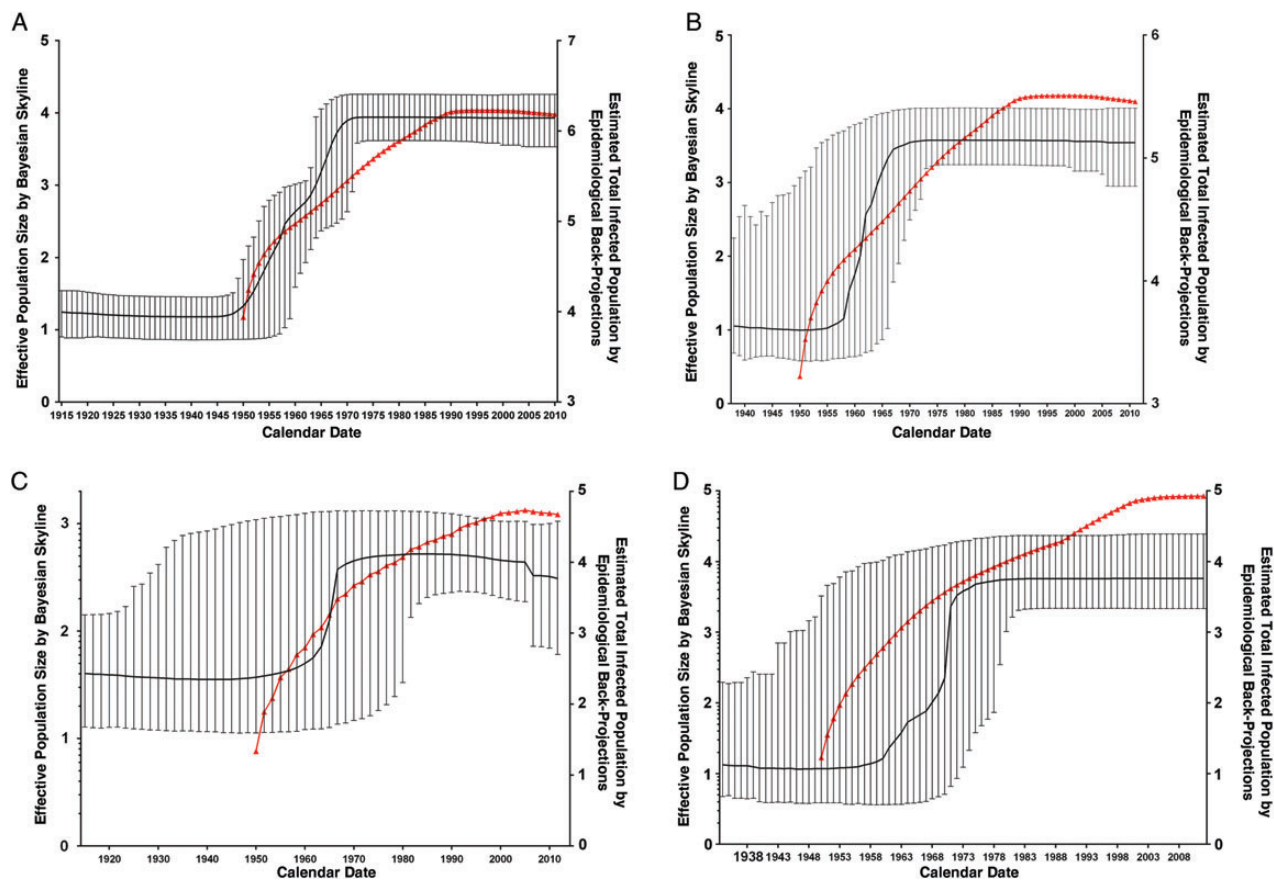
Figure 1. Bayesian skyline simulations of the effective population size with hepatitis C virus infections in North America (*A, B*) and Australia (*C, D*) compared with epidemiological back-projections. Solid line represent change in median effective population size (left y-axis) with time (x-axis, calendar years), as estimated by the Bayesian evolutionary analysis. Effective population size refers to a population inferred from the available sequence data and not to real incidence or prevalence values in the population. Error bars represent 95% high posterior density (HPD) intervals for each time point. Lines with triangles represent total infected population as estimated by epidemiological back-projections (right y-axis) using real prevalence estimates for the population under study. Both y-axes are on a logarithmic scale. The epidemiological projections are available only up to 1950, whereas the Bayesian projections extend to the lower HPD interval of the estimated time of the most recent common ancestor (Table 2). Only the trends of y-axes may be compared, not the absolute numbers, owing to the intrinsic differences in methods and output parameters mentioned above. Panels show subtypes 1a (*A, C*) and 3a (*B, D*).

era. The maximum clade credibility trees for each data set are given in the Supplementary material (S2).

Comparison of Coalescent Projections With Epidemiological Model Projections

The BEAST skyline output for each data set was superimposed on data derived from the epidemiological model estimates derived using the previously published system dynamic modeling framework [20, 21]. Comparison of the trends reveals considerable similarities with regard to the timing of the exponential increases in the effective population size with the Bayesian phylogenetic analysis and the size of the HCV-infected population from epidemiological back-projections for both genotypes and continents (Figure 1).

DISCUSSION

These convergent virological and epidemiological analyses have allowed tracing of historical trends in the cross-continental spread of HCV subtype 1a and 3a epidemics in North America and Australia, and identified dramatic increases in the relative size of the epidemics in the mid-20th century. Because the HCV genome features widely varied evolutionary rates, within infected hosts, across infected populations, and between different regions within the genome, assumptions regarding these rates have the potential to influence the outcome of the complex statistical methods applied in the coalescent analysis described here. Accordingly, for the present analysis, the substitution rates for the full genome estimated by Gray et al [17] were used as priors. It should be noted that because most sequences used in our study were from acute infections, they could plausibly have a higher substitution rate in the envelope region, given that the virus is recognized to mutate frequently in this region as it adapts for efficient entry and to the immune responses within the new host. However, because the full-length genome was used, the averaged substitution rate across the genome is

not likely to vary significantly from the estimates used in the previous study [17].

The tMRCA estimates for both North American and Australian 1a sequences were in the early 20th century, suggesting the virus was present in both continents at the beginning of the century. Previous studies have estimated the tMRCA for subtype 1a in other parts of the world, using limited segments of the genome. For instance, a previous global evolutionary analysis of subtype 1a, using the NS5B region of sequences obtained from 21 countries, placed the tMRCA at a similar time point and with a Bayesian skyline plot showing a similar pattern to that reported here for North America [1]. However, this study included a disproportionate number of North American sequences (399 vs 443 sequences from 20 other countries combined). Regarding regional studies, using core-E1 sequences from Vietnam, Li et al [27] placed the tMRCA of subtype 1a around 1928 in Southeast Asia. By contrast, 2 studies from South America using 1a NS5B sequences placed the tMRCA for Brazilian sequences between 1950 and 1979. Even after adjusting for the slower rate of evolution of NS5B, this estimate is more recent than estimates of the origin of the virus in other continents. Taken together, these data suggest that in Australia, North America, and Southeast Asia, subtype 1a originated at the beginning of the 20th century, with a more recent introduction into South America.

The epidemic of genotype 3a has also been examined in previous studies [4, 15, 28, 29], including 2 that used full-length genomes [15, 29]. However, for both of these previous studies, <50 full-length 3a sequences were available from across the globe with very few or no sequences from United States and Australia. With the InC3 sequencing project, the number of full-length subtype 3a sequences available has been more than doubled (67 new sequences added). The current data set now allows tracing of the global spread of subtype 3a infections by provision of previously missing information from North America and Australia and has placed the tMRCA at about 1950 in both continents. Li et al [30] used 42 sequences of 8 distinct genotype 3 subtypes and estimated the tMRCA to be around 1934 (95% HPD, 1915–1949) for subtype 3a. Choudhary et al [15] also constructed a time-scaled phylogeny of 55 subtype 3a full-length sequences obtained from public databases with a global distribution and estimated the tMRCA to be approximately 1913 (95% HPD, 1859–1944). Interestingly, in both of these previous studies, the first sequences to branch out from the root were from India. In the latter study, the only sequence from Australia, and 3 sequences from United States, all branched out at about 1950, as did other sequences from Denmark, Italy, China and New Zealand. Two other studies from South America that used partial 3a sequences from Brazil, placed the tMRCA around 1967–1979, which is considerably more recent [4, 28].

Comparison of the trends predicted by the coalescent analysis of phylogenetic data and the epidemiological model estimates were remarkably similar with regard to the rapid rise in infections from the late 1950s onward and the later stabilization of the epidemics, which serves as validation for the predictions by 2 entirely different methods. It should be noted, however, that both data sets are limited by the first available raw data points (1977 and 1989 for the 1a and 3a viral sequences, respectively, and 1997 for the epidemiological data). Having more sequences spread across the time period of interest allows better estimation of parameters including substitution rate. For example, the subtype 1a analysis in Australia, where the distance between the predicted root and the oldest available sequence is the largest, the HPD intervals around tMRCA are also large. The robustness of coalescent analyses in predicting HCV epidemics has already been established [18], and a recent study using this method on subtype 1a data in North America yielded a very similar prediction to that reported here regarding the timing of the epidemic [31]. Regarding the epidemiological modeling, the data are dependent on reported numbers in national surveys and notification systems, which may be affected by underreporting.

The genetic diversity between HCV subtypes is immense, with some subtypes thought to have existed much earlier than others (as endemic infections in restricted geographic locations). For example, the origins of subtype 6 in East Asia have been traced back to approximately 1000 CE by evolutionary analysis of sequence data [32]. Similarly, the origin of the genotype 4 epidemic in Africa has been traced back to the 17th century [33]. However, evolutionary analysis of the more common subtypes with a global distribution (1a, 1b, and 3a), which are frequently referred to as "epidemic" subtypes, have consistently traced the ancestor to within the last century in many locations around the world [34, 35]. The original endemic viral population sources for these emergent subtypes have not been identified.

The origins of subtypes 1a and 3a in North America and Australia generally coincide with the First and Second World Wars (WWI and WWII), respectively, and are likely to be closely linked to developments in medical technology and practice in that era. The United States mobilized >1 million soldiers to battlefronts in Europe in WWI and 16 million in WWII [36]. There were 330 000 Australians who enlisted for military service and fought in Europe in WWI, constituting approximately 7% of the total population at that time [37], and 1 million served abroad in WWII [37, 38]. Similarly, the Korean War (1950–1953) allowed the mixing of large numbers of young Australians and Americans [39, 40]. In conjunction with these conflicts blood transfusion for management of major trauma was first undertaken in 1916 [41], followed by establishment of the first blood donation service [41]. In the same time period, there was rapid growth in the use of hypodermic needles and syringes initially for subcutaneous, and subsequently intravenous, administration of

medications. There was a dramatic fall in prices for these devices between 1920 and 1950 [42]. In addition, there had been a preceding rise in the routine administration of opium and morphine by physicians, which is cited as the leading cause of opiate dependency at that time [43].

The Bayesian skyline plots for subtypes 1a and 3a reveal a period of exponential growth of infections in both North America and Australia between 1955 and 1975. Soon after WWII, injecting drug use became popular in United States and was accompanied by a post-War surge in heroin use. Armstrong et al [22, 23] modeled the lifetime probability of injecting drug use and the prevalence of anti-HCV positivity by birth year for the United States. This analysis shows that, in both cases, numbers reached a peak for cohorts born in 1948–1962. The age at initiation of injecting drug use is predominantly in the late teens or early 20s [44], which suggests that the peak HCV prevalence may follow 15–20 years later. This correlates well with the Bayesian skyline projections described here (Supplementary material S3). In Australia, there are no data on drug use prior to 1960 [45]. Recent models estimating injecting drug use patterns in Australia assume that the prevalence increased by approximately 8% per year between 1970 and 1997 [45]. The simulated rise in the number of HCV-infected individuals for both subtypes in North America and Australia identified here thus correlates with the surge in injecting drug use.

In both continents since the mid-1970s, the projections show stabilization of the epidemic. Since the early 1980s, the risks of blood-borne virus transmission, notably for HIV and subsequently HCV, have become well known to the public and led to harm-minimization strategies for injecting drug users. In the United States, the incidence of HIV infections in persons who inject drugs has dropped by 50% from 1997–2009 (80% since 1980) [46]. Surveillance data indicate that the incidence of acute hepatitis C also declined from 2000 to 2004 and remained stable until 2010, after which the incidence has been on an increasing trend, probably owing to increasing injecting drug use [47]. The findings of the current study are consistent with these observations, although they failed to show the later surge of infections, because the most recent samples were collected in 2011 in the United States. In Australia, epidemiological estimates and surveillance data suggest that injecting drug use and incident HCV infections have declined since the early 2000s [45, 48]. There is a stable or falling trend in the prevalence of hepatitis C among participants in needle and syringe programs, prison entrants, and blood donors [49, 50]. These data are consistent with the trends in the projections described here.

In conclusion, using coalescent analysis of full-length HCV genomes, the historical trends in the HCV epidemics in Australia and North America have been mapped with convergent data derived from epidemiological modeling. The subtype 3a epidemic has more recent origins than the subtype 1a epidemic. All projections showed an exponential increase in infections from 1955 to 1975, coincident with the increase in injecting drug use. The epidemics subsequently stabilized up to the time window examined, potentially explained by improved harm-minimization measures after the discovery of HIV and HCV.

Supplementary Data

Notes

References

1. Magiorkinis G, Magiorkinis E, Paraskevis D, et al. The global spread of hepatitis C virus 1a and 1b: a phylodynamic and phylogeographic analysis. PLoS Med 2009; 6: e1000198.
2. de Bruijne J, Schinkel J, Prins M, et al. Emergence of hepatitis C virus genotype 4: phylogenetic analysis reveals three distinct epidemiological profiles. J Clin Microbiol 2009; 47:3832–8.
3. Sunbul M, Khan A, Kurbanov F, et al. Tracing the spread of hepatitis C virus in Turkey: a phylogenetic analysis. Intervirology 2013; 56:201–5.
4. Lampe E, Espirito-Santo MP, Martins RM, Bello G. Epidemic history of hepatitis C virus in Brazil. Infect Genet Evol 2010; 10:886–95.
5. Hajarizadeh B, Grebely J, Dore GJ. Epidemiology and natural history of HCV infection. Nat Rev Gastroenterol Hepatol 2013; 10:553–62.
6. Smith DB, Bukh J, Kuiken C, et al. Expanded classification of hepatitis C virus into 7 genotypes and 67 subtypes: updated criteria and genotype assignment web resource. Hepatology 2014; 59:318–27.
7. Messina JP, Humphreys I, Flaxman A, et al. Global distribution and prevalence of hepatitis C virus genotypes. Hepatology 2015; 61:77–87.
8. Gire SK, Goba A, Andersen KG, et al. Genomic surveillance elucidates Ebola virus origin and transmission during the 2014 outbreak. Science 2014; 345:1369–72.
9. Rodrigo A. The coalescent: population genetic inference using genealogies. In: Lemey P, Salemi M, Vandamme A, eds. The phylogenetic handbook: a practical approach to phylogenetic analysis and hypothesis testing. 2nd ed. Cambridge, United Kingdom: Cambridge University Press, 2009:551–61.

10. Rodrigo C, Eltahla AA, Bull R, et al. Phylogenetic analysis of full-length, early infection, hepatitis C virus genomes among people with intravenous drug use: the InC3 Study. J Viral Hepat **2016**. In Press.

11. Grebely J, Morris MD, Rice TM, et al. Cohort profile: the International Collaboration of Incident HIV and Hepatitis C in Injecting Cohorts (InC³) study. Int J Epidemiol **2013**; 42:1649–59.

12. Bull RA, Eltahla AA, Rodrigo C, et al. A method for near full-length amplification and sequencing for six hepatitis C virus genotypes. BMC Genomics **2016**; 17:247.

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

14. Kumar S, Nei M, Dudley J, Tamura K. MEGA: a biologist-centric software for evolutionary analysis of DNA and protein sequences. Brief Bioinform **2008**; 9:299–306.

15. Choudhary MC, Natarajan V, Pandey P, et al. Identification of Indian sub-continent as hotspot for HCV genotype 3a origin by Bayesian evolutionary reconstruction. Infect Genet Evol **2014**; 28:87–94.

16. Yuan M, Lu T, Li C, Lu L. The evolutionary rates of HCV estimated with subtype 1a and 1b sequences over the ORF length and in different genomic regions. PLoS One **2013**; 8:e64698.

17. Gray RR, Parker J, Lemey P, Salemi M, Katzourakis A, Pybus OG. The mode and tempo of hepatitis C virus evolution within and among hosts. BMC Evol Biol **2011**; 11:131.

18. Drummond AJ, Rambaut A, Shapiro B, Pybus OG. Bayesian coalescent inference of past population dynamics from molecular sequences. Mol Biol Evol **2005**; 22:1185–92.

19. Rambaut A, Suchard MA, Xie D, Drummond AJ. Tracer v1.6. http://beast.bio.ed.ac.uk/Tracer. Accessed 1 March 2016.

20. Razavi H, Elkhoury AC, Elbasha E, et al. Chronic hepatitis C virus (HCV) disease burden and cost in the United States. Hepatology **2013**; 57:2164–70.

21. Razavi H, Waked I, Sarrazin C, et al. The present and future disease burden of hepatitis C virus (HCV) infection with today's treatment paradigm. J Viral Hepat **2014**; 21(suppl 1):34–59.

22. Armstrong GL. Injection drug users in the United States, 1979–2002: an aging population. Arch Intern Med **2007**; 167:166–73.

23. Armstrong GL, Wasley A, Simard EP, McQuillan GM, Kuhnert WL, Alter MJ. The prevalence of hepatitis C virus infection in the United States, 1999 through 2002. Ann Intern Med **2006**; 144:705–14.

24. Denniston MM, Jiles RB, Drobeniuc J, et al. Chronic hepatitis C virus infection in the United States, National Health and Nutrition Examination Survey 2003 to 2010. Ann Intern Med **2014**; 160:293–300.

25. Myers RP, Krajden M, Bilodeau M, et al. Burden of disease and cost of chronic hepatitis C infection in Canada. Can J Gastroenterol Hepatol **2014**; 28:243–50.

26. Sievert W, Razavi H, Estes C, et al. Enhanced antiviral treatment efficacy and uptake in preventing the rising burden of hepatitis C-related liver disease and costs in Australia. J Gastroenterol Hepatol **2014**; 29(suppl 1):1–9.

27. Li C, Yuan M, Lu L, et al. The genetic diversity and evolutionary history of hepatitis C virus in Vietnam. Virology **2014**; 468–470:197–206.

28. Nishiya AS, de Almeida-Neto C, Romano CM, et al. Phylogenetic analysis of the emergence of main hepatitis C virus subtypes in Sao Paulo, Brazil. Braz J Infect Dis **2015**; 19:473–8.

29. Bility MT, Cheng L, Zhang Z, et al. Hepatitis B virus infection and immunopathogenesis in a humanized mouse model: induction of human-specific liver fibrosis and M2-like macrophages. PLoS Pathog **2014**; 10:e1004032.

30. Li C, Lu L, Murphy DG, Negro F, Okamoto H. Origin of hepatitis C virus genotype 3 in Africa as estimated through an evolutionary analysis of the full-length genomes of nine subtypes, including the newly sequenced 3d and 3e. J Gen Virol **2014**; 95:1677–88.

31. Joy JB, McCloskey RM, Nguyen T, et al. The spread of hepatitis C virus genotype 1a in North America: a retrospective phylogenetic study. Lancet Infect Dis **2016**; 16:698–702.

32. Pybus OG, Barnes E, Taggart R, et al. Genetic history of hepatitis C virus in East Asia. J Virol **2009**; 83:1071–82.

33. Iles JC, Raghwani J, Harrison GL, et al. Phylogeography and epidemic history of hepatitis C virus genotype 4 in Africa. Virology **2014**; 464–465:233–43.

34. Pybus OG, Charleston MA, Gupta S, Rambaut A, Holmes EC, Harvey PH. The epidemic behavior of the hepatitis C virus. Science **2001**; 292:2323–5.

35. Tanaka Y, Hanada K, Mizokami M, 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**; 99:15584–9.

36. US Army Europe. History. http://www.eur.army.mil/organization/history.htm. Accessed March 25 2016.

37. Australian Government. National Archives of Australia: Boer War to Vietnam. http://www.naa.gov.au/collection/explore/defence/conflicts.aspx. Accessed 25 March 2016.

38. WWII Foundation. WWII facts & figures. http://www.wwiifoundation.org/students/wwii-facts-figures/. Accessed 25 March 2016.

39. McCoy AW. Drug traffic: narcotics and organized crime in Australia. Sydney, Australia: Harper and Row, **1980**.

40. Robins LN, Helzer JE, Davis DH. Narcotic use in Southeast Asia and afterward: an interview study of 898 Vietnam returnees. Arch Gen Psychiatry **1975**; 32:955–61.

41. Giangrande PL. The history of blood transfusion. Br J Haematol **2000**; 110: 758–67.

42. Drucker E, Alcabes PG, Marx PA. The injection century: massive unsterile injections and the emergence of human pathogens. Lancet **2001**; 358:1989–92.

43. Courtwright D. Dark paradise: opiate addiction in America before 1940. Boston, MA: Harvard University Press, **1982**: 42.

44. National Institute on Drug Abuse. Drug facts: nationwide trends. https://www.drugabuse.gov/publications/drugfacts/nationwide-trends. Accessed 12 February 2016.

45. Razali K, Amin J, Dore GJ, Law MG; HCV Projections Working Group. Modelling and calibration of the hepatitis C epidemic in Australia. Stat Methods Med Res **2009**; 18:253–70.

46. HIV-associated behaviors among injecting-drug users—23 cities, United States, May 2005-February 2006. MMWR Morb Mortal Wkly Rep **2009**; 58:329–32.

47. Centers for Disease Control and Prevention. Surveillance for viral hepatitis—United States, 2013. http://www.cdc.gov/hepatitis/statistics/2013surveillance/index.htm#tabs-801919-8. Accessed 2 November 2015.

48. Iversen J, Wand H, Topp L, Kaldor J, Maher L. Reduction in HCV incidence among injection drug users attending needle and syringe programs in Australia: a linkage study. Am J Public Health **2013**; 103:1436–44.

49. Australian Department of Health. Fourth National Hepatitis C Strategy 2014–2017. Canberra, Australia: Department of Health, **2014**.

50. Australian Department of Health. Seventh National HIV Strategy. Canberra, Australia: Department of Health, **2014**.