

The hepatitis C virus epidemic among injecting drug users

Oliver G. Pybus^{a,*}, Alexandra Cochrane^{b,1},
Edward C. Holmes^a, Peter Simmonds^b

^aDepartment of Zoology, University of Oxford, South Parks Road, Oxford, UK

^bLaboratory for Clinical and Molecular Virology, University of Edinburgh, Edinburgh, UK

Received 28 June 2004; received in revised form 4 August 2004; accepted 4 August 2004

Available online 11 September 2004

Abstract

Given the economic and health costs of hepatitis C virus (HCV) infection, and the ongoing transmission within the injecting drug user (IDU) population, there is a need for improved understanding of HCV epidemiology within this risk group. We employed a recently developed method based on phylogenetic analysis to infer HCV epidemic history and to provide the first estimates of the rate of spread of subtypes 1a and 3a circulating within injecting drug user populations. The data indicates that HCV subtype 1a entered the IDU population on at least three separate occasions. Both subtypes demonstrate exponential population growth during the 20th century, with a doubling time of 7–8 years. The results provide a baseline for prediction of the future course of the HCV epidemic, and its likely response to transmission control policies. © 2004 Elsevier B.V. All rights reserved.

Keywords: Molecular epidemiology; Public health; Injecting drug use; Evolution; Phylogeny; HCV

1. Introduction

Within industrialised nations the transmission of hepatitis C virus (HCV) through blood products has effectively been halted (Schreiber et al., 1996), leaving injecting drug use as the major risk activity (Trepo and Pradat, 1999). Although HCV spread within this risk group has slowed following the introduction of needle exchange and educational programs (Goldberg et al., 2001; Taylor et al., 2000) the transmission rate remains high (Crofts et al., 1999). At least 50% of people infected with HCV develop chronic hepatitis with attendant morbidity, mortality and an estimated annual cost of \$5.46 billion in the USA alone (Leigh et al., 2001). Public health goals therefore include prediction of the future course of the epidemic to allow appropriate resource allocation and development of an effective transmission control policy. These goals should be grounded in a clear understanding of the transmission dynamics of the current HCV epidemic in the injecting drug user (IDU) risk group. Available evidence

from molecular epidemiology (Smith et al., 1997) suggests longstanding endemic HCV infection of populations in Africa and Asia, with movement into new transmission networks in industrialised nations, resulting in the recent epidemic in both transfusion recipients and IDUs (Fig. 1). Age-stratified seroprevalence data (Alter et al., 2001) and phylogenetic analysis (Pybus et al., 2001) indicate a rapid recent spread of HCV in industrialised nations, but to date no risk group-specific estimates of the rate of spread or the shape of the epidemic curve have been obtained.

Typically, the rate of spread of an infectious agent through a population is derived from infection incidence data collected over the time-course of the epidemic. Such data is difficult to obtain for the current HCV epidemic for a number of reasons. First, the beginning of the epidemic predated the discovery of the virus, and the lack of suitable archived specimens over 30 years old hinders retrospective seroprevalence measurement. Second, the absence of specific symptoms of either initial infection or chronic disease makes it difficult to estimate past prevalence from medical records.

In view of these difficulties, we have used a recently developed molecular epidemiological method to investigate

* Corresponding author. Tel.: +44 1865 271274; fax: +44 1865 271249.

E-mail address: oliver.pybus@zoo.ox.ac.uk (O.G. Pybus).

¹ These authors contributed equally.

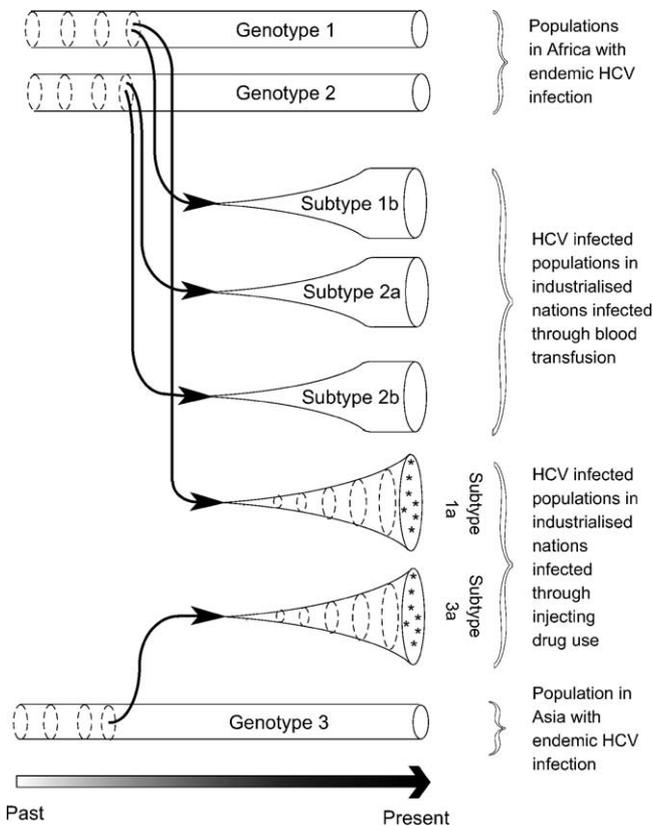


Fig. 1. Simplified representation of the history of the current HCV epidemic, illustrating the population studied in this paper. The cylinders and cones represent populations of HCV-infected individuals through time (cylinders represent constant sized populations and cones represent growing populations). Cross-sections (dashed circles) represent the changing size of the ancestral populations. The curved arrows indicate the seeding of different HCV strains into new transmission networks within industrialised nations. In the method used here, HCV is sampled from HCV-infected IDUs (indicated by asterisks). Through phylogenetic analysis we are able to infer the size of the ancestral population over time. Note that the method has the potential to look back into the endemic populations, prior to the seeding of the IDU epidemic.

the epidemic history of HCV in the IDU population. The method, based on coalescent theory (see Donnelly and Tavare, 1995), estimates the past size of an infected population by phylogenetic analysis of viral gene sequences sampled from the infected population. In other words, through analysis of HCV sequences amplified from currently infected IDUs, we are able to describe the epidemic history of HCV infection in this risk group (Fig. 1). The coalescent approach is established as a useful tool for reconstructing the past population dynamics of human pathogens, including HCV (Pybus et al., 2001, 2003; Tanaka et al., 2002, 2004; Nakano et al., 2004), HIV (Pybus et al., 2000; Strimmer and Pybus, 2001; Lemey et al., 2003) and malaria (Joy et al., 2003).

The change in the estimated number of HCV infections through time has been used to calculate the growth rate and the basic reproductive number, R_0 , of HCV (Pybus et al., 2001). R_0 is the number of secondary infections generated by a single infectious individual at the beginning of an

epidemic. Where an epidemic is initiated by the introduction of an infection into a large susceptible population, R_0 reflects the intrinsic replicative capacity of the infection (Anderson and May, 1991). However, the models used to estimate R_0 from genetic data assume that the host population size remains constant through time (Pybus et al., 2001). As discussed later, this is not the case for the IDU epidemic studied here, since the rate of increase in the number of infections is likely to have followed the increasing size of the IDU population. Thus, the “ R_0 values” we obtain most closely represent the growth in size of the IDU risk group, rather than the intrinsic transmission capacity of the virus itself. This result provides an important new perspective on the analysis of pathogen genetic diversity using coalescent theory.

Two sequence datasets were analysed in this study. The first dataset, hereafter termed the UK dataset, comprised sequences amplified from 90 IDUs attending health services in three cities in the UK and from 24 blood donors donating in Scotland. Analysis of this dataset provides an estimate of the epidemic history of HCV circulating within the UK IDU risk group. The second dataset, termed the extended dataset, comprised the UK sequences plus sequences obtained from IDUs in Australia and France.

2. Methods

2.1. Samples

HCV subtype 1a and 3a NS5B consensus sequences were amplified from a total of 146 serum samples, drawn between 1997 and 2001 from 146 HCV-infected IDUs attending health services in London, Glasgow, Edinburgh, Marseilles or Melbourne (Cochrane et al., 2002). The number of sequences included from each city is given in Table 1. The mean year of sampling was 2000. The majority of samples were from individuals either aged less than 30, or with duration of injecting activity of less than 15 years. In addition, 12 sequences of each subtype were obtained from the stored serum of prospective blood donors in Scotland. The study protocol conformed to the ethical guidelines of the Lothian Research Ethics Committee.

2.2. Laboratory methods

The laboratory methods used to obtain the HCV NS5B consensus sequences have been fully described previously

Table 1
Number of sequences obtained from each city

HCV subtype	1a	3a
London	11	16
Glasgow	14	21
Edinburgh	15	15
Marseilles	19	14
Melbourne	9	12

(Cochrane et al., 2002). In brief, total RNA was extracted from serum using phenol chloroform, and two overlapping fragments were reverse transcribed and amplified using the ACCESS RT-PCR kit (Promega Corporation, Madison, USA) followed by nested PCR. The amplified fragments were sequenced in both sense and antisense directions, edited by hand, and the portion from nucleotide 8094 to 8788 (subtype 1a) and from 8094 to 8777 (subtype 3a) was used for phylogenetic estimation (numbering according to HPCPLYPRE, accession number M62321). Sequences are submitted to GenBank under accession numbers AY100024–AY100193.

2.3. Estimation of epidemic history and demographic parameters

We use a statistical method, based on coalescent theory, to estimate the epidemic history of HCV in IDUs from the viral gene sequences obtained. A coalescent process is a population genetic model that describes the relationship between the demographic history of a population and the genealogy of individuals sampled randomly from it. In short, the shape of the genealogy can be used to estimate past numbers of infections in the sampled population. A full mathematical description of the method is not possible here; for more details see (Holmes et al., 1995; Pybus et al., 2000; Pybus and Rambaut, 2002). The coalescent approach can be broken down into the following steps:

- (i) HCV strains from the population under investigation are sampled and sequenced.
- (ii) The genealogy of the HCV strains obtained in step (i) is estimated. Standard phylogenetic methods are employed; we used the program PAUP* (Swofford, 1998).
- (iii) The timescale of the genealogy is converted. The genealogy from step (ii) is measured in units of genetic distance. To convert the genealogy into a timescale of years, we estimate the rate of molecular evolution of the sampled sequences. The program TipDate (Rambaut, 2000) was used to perform this step.
- (iv) The epidemic history and epidemiological parameters are estimated from the converted genealogy. This calculation uses coalescent theory and was performed by the program GENIE (Pybus and Rambaut, 2002).

In the following sections, we describe the methods used in steps (ii), (iii) and (iv) in detail.

2.3.1. Estimation of the genealogies of the sampled HCV strains

HCV subtype 1a and 3a genealogies were estimated from the UK and extended NS5B sequence datasets using a maximum likelihood approach, as implemented in PAUP* (Swofford, 1998). For each dataset, a heuristic search was employed to identify the genealogy that maximises the log

likelihood of the sequence data, given a model of sequence evolution. The general time-reversible (GTR) model was used, with a codon-position model of among-site rate variation and with the molecular clock enforced. The molecular clock was tested using a likelihood ratio test (Felsenstein, 1981) and was accepted for subtype 3a but not subtype 1a. This suggests there is some rate heterogeneity among lineages within HCV subtypes, but that this variation is most likely too small to significantly bias our estimates (see Section 4 for more consideration of this point).

2.3.2. Converting evolutionary distance into time

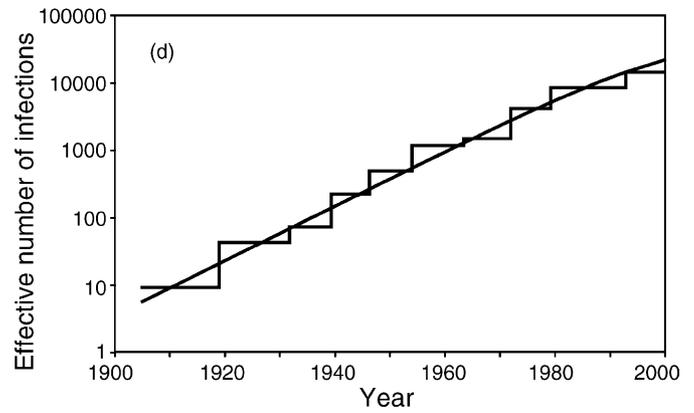
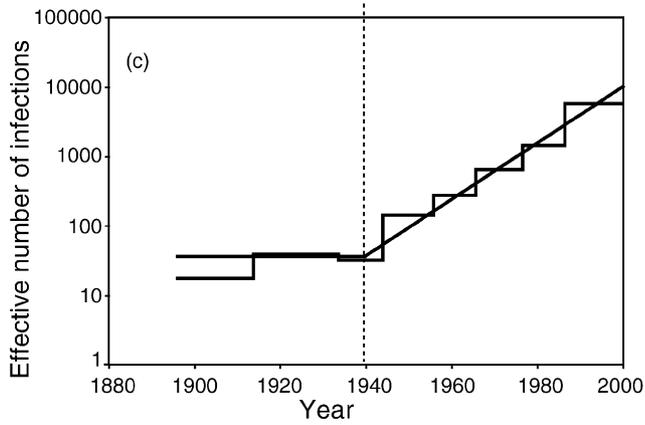
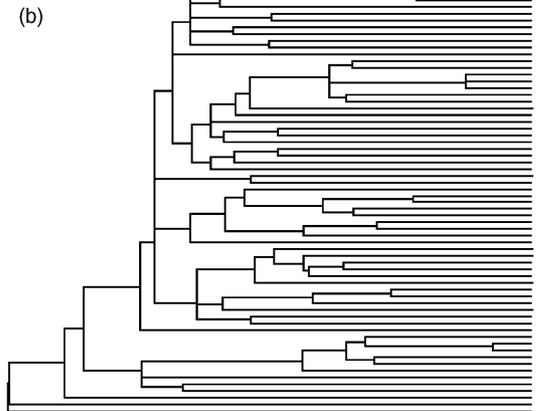
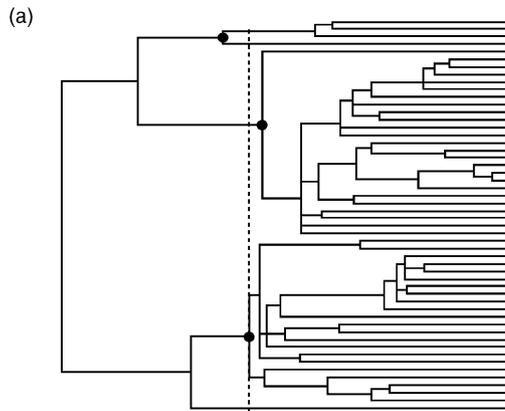
Evolutionary distance was converted into time using a evolutionary rate estimated from HCV subtype 1b sequences sampled from fourteen women, all of whom had acquired the infection 17 years earlier through exposure to a contaminated batch of anti-rhesus D immunoglobulin (Power et al., 1995). (The applicability of this rate to subtypes 1a and 3a is discussed in Section 4.) As evolutionary rates significantly differ between HCV genome regions, the same 695 base pair segment of NS5B used to construct the IDU genealogies was used to estimate evolutionary rate (accession numbers AY321564–AY321577). As previously described, the mean rate of nucleotide substitution was calculated from the sequences using the program TipDate (Rambaut, 2000) and a star-shaped phylogeny. The estimated rate was 4×10^{-4} substitutions per site per year (95% confidence intervals 3×10^{-4} to 5×10^{-4}).

2.3.3. Estimation of epidemic histories and demographic parameters

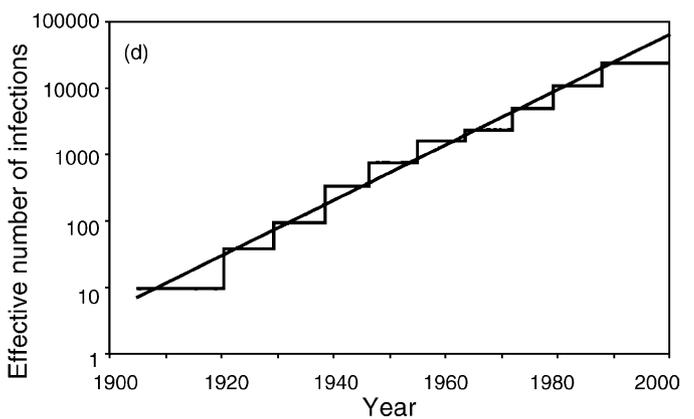
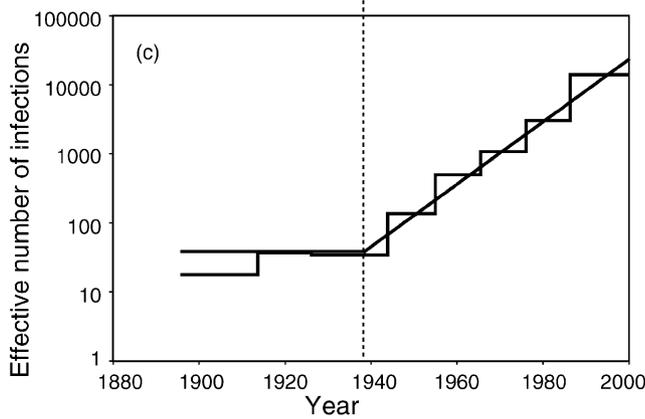
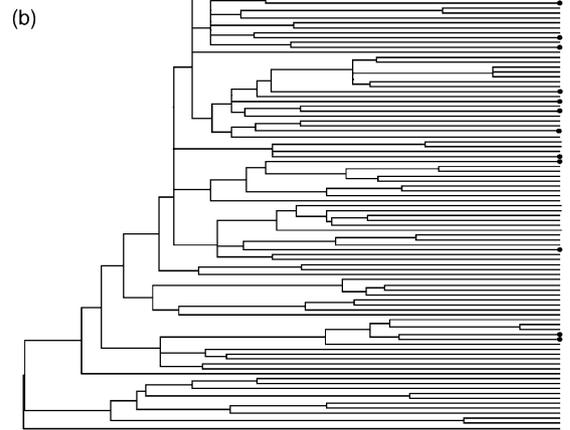
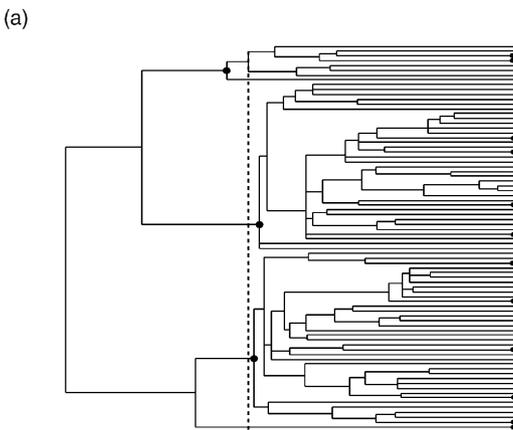
HCV epidemic history was estimated using non-parametric and parametric methods based on coalescent theory. Under this theory, the shape of the genealogy of a sample of viral gene sequences may be used to infer the demographic history of that population (Donnelly and Tavaré, 1995). The coalescent framework assumes that the shape of the genealogy is largely unaffected by recombination, selection and population subdivision, that genetic distance is proportional to time, and that sample size is considerably smaller than effective population size. See Section 4 and references for further consideration of these assumptions.

The results for each dataset are presented as plots of the effective number of HCV infections through time. The effective number of infections at time t is considered proportional to the number of HCV infections in the sampled population at that time, with the constant of proportionality being a function of the ‘generation time’ of HCV infections and the variance among infected individuals in onward virus transmission (Donnelly and Tavaré, 1995). On each plot, we superimpose parametric and non-parametric estimates of epidemic history. The parametric curve is estimated from the HCV genealogy using maximum likelihood under a demographic model (e.g. exponential growth, logistic

(2)



(3)



growth), and provides estimates of parameters such as r , the rate of exponential population growth. The relative goodness-of-fit of all nested demographic models was assessed using the likelihood ratio statistic (LRS). Details of the demographic models used in this analysis and their associated parameters are provided in the Appendix. Following Pybus et al. (2001), R_0 is calculated from r using the equation $R_0 = rD + 1$, where D is the average duration of infectiousness (see Section 4 for the correct interpretation of these values).

The non-parametric estimate, called the skyline plot (Strimmer and Pybus, 2001), does not assume a specific demographic model and is therefore used as a model selection tool, indicating whether the parametric curve provides a good fit to the data or not.

3. Results

HCV subtype 1a and 3a phylogenies were reconstructed from the UK dataset (Fig. 2a and b) and from the extended dataset (Fig. 3a and b). The average pairwise genetic distance between sequences is approximately 0.08 substitutions per site. The sequences amplified from the blood donors were scattered amongst the sequences amplified from IDUs (Fig. 3a and b) suggesting circulation of the virus between blood donors and the IDU risk group. Sequences from the blood donors were thus included in subsequent analyses. As demonstrated previously (Cochrane et al., 2002), sequences amplified from IDUs in Marseilles and Melbourne formed phylogenetic clusters, indicating partial segregation between HCV circulating within these cities from HCV circulating within the UK.

Using the estimated HCV genealogies, the epidemic histories of the sampled populations were inferred. Estimates of the effective number of HCV infections through time are shown in Fig. 2c and d (UK dataset) and in Fig. 3c and d (extended dataset). The estimates represent epidemic history from the time of divergence of the sampled viruses (approximately 1900) to the time of sampling (year 2000). In both cases, the plot for subtype 1a best fits the piecewise expansion demographic model (LRS = 4.26, $p < 0.05$ for the UK dataset; LRS = 10.26, $p < 0.01$ for the extended dataset). This is a two-phase model of population growth, consisting of an initial period of constant population size followed by a period of exponential growth (see Appendix). The onset of exponential growth begins around 1940 (95% confidence limits

1924–1953) and coincides with the appearance of three clades in the subtype 1a phylogenetic tree (Figs. 2a and 3a). The HCV genotype 3a datasets best fit a model of logistic growth, although the recent reduction in growth rate observed is not statistically significant and a simpler model of exponential growth cannot be rejected (LRS = 0.12, $p > 0.05$ for the UK dataset; LRS = 0.46, $p > 0.05$ for the extended dataset).

Estimates of the epidemiological parameters r (the growth rate achieved during the exponential growth phase) and R_0 (the basic reproductive number) are given for each sequence dataset in Table 2. R_0 was calculated using a range of four plausible durations of infectiousness (D).

4. Discussion

The epidemic histories presented here are the first to be estimated for HCV circulating within the IDU population. One hundred and forty-six of the 170 sequences used in this study were obtained from documented injecting drug users, providing a large IDU-specific HCV sequence dataset. The additional 24 sequences from blood donors were phylogenetically interspersed with the IDU sequences, suggesting mixing of the virus between the IDU and blood donor risk groups. The estimated parameters are average values for the whole study population and local differences in needle sharing behaviour may lead to divergence from the values quoted.

The estimated epidemic histories indicate a period of exponential growth during the last century for both HCV subtype 1a and 3a. This is in accordance with data from the USA, where a rapid increase in HCV infection incidence during the second half of the 20th century (as estimated from age-specific seroprevalence data) is thought to be due to transmission in the IDU risk group (Alter et al., 2001). Previous estimates of HCV epidemic history, obtained using the coalescent framework employed here, have also indicated recent exponential growth for subtypes 1a and 1b globally, and for subtype 4a in Egypt (Pybus et al., 2001, 2003; Tanaka et al., 2002, 2004; Nakano et al., 2004), again in accordance with current theories of HCV spread.

The two-phase epidemic history presented for HCV subtype 1a is in keeping with previous studies suggesting that the global subtype 1a epidemic originated from an ancestral population in West Africa (Jeannel et al., 1998). Thus the constant population size phase possibly reflects

Figs. 2–3. Genealogies and corresponding estimates of epidemic history for HCV subtypes 1a and 3a, drawn on the same timescale. Fig. 2 shows the UK dataset, while Fig. 3 shows the extended dataset. Parts a and b: genealogies estimated from the subtype 1a (a) and subtype 3a (b) gene sequences. The positions of sequences obtained from blood donors are indicated on the Fig. 3 genealogies with a small circle. The dotted line represents the estimated year of change from constant population size to exponential growth and the large circles represent proposed introductions of endemic strains into new transmission networks. Parts c and d: Estimates of epidemic history for subtype 1a (c) and subtype 3a (d), drawn on the same time scale as the corresponding genealogies. The black lines represent parametric estimates obtained using the piecewise expansion (c) and logistic (d) models (see Section 2 and Appendix). The grey lines represent corresponding non-parametric estimates (skyline plots; see Section 2). The effective number of infections is equivalent to the inbreeding effective population size of the epidemic.

Table 2
Estimates of epidemiological parameters for HCV subtypes 1a and 3a

Dataset	Number of sequences	Exponential growth rate ^a , r (95% confidence limits)	D^b	Basic reproductive number, R_0^c (95% confidence limits)
Subtype 1a, UK dataset	52	0.092 (0.069, 0.11)	10	1.92 (1.69, 2.10)
			15	2.38 (2.04, 2.65)
			20	2.84 (2.38, 3.2)
			25	3.30 (2.73, 3.75)
Subtype 3a, UK dataset	64	0.093 (0.077, 0.11)	10	1.93 (1.77, 2.10)
			15	2.40 (2.15, 2.65)
			20	2.86 (2.54, 3.2)
			25	3.33 (2.93, 3.75)
Subtype 1a, extended dataset	80	0.104 (0.089, 0.127)	10	2.04 (1.89, 2.27)
			15	2.56 (2.34, 2.91)
			20	3.08 (2.78, 3.54)
			25	3.60 (3.23, 4.17)
Subtype 3a, extended dataset	90	0.096 (0.083, 0.108)	10	1.96 (1.83, 2.08)
			15	2.44 (2.25, 2.62)
			20	2.92 (2.66, 3.16)
			25	3.40 (3.08, 3.70)

^a Units of r are year⁻¹.

^b D is the average duration of infectiousness in years.

^c See Section 4 for detailed interpretation of these values.

endemic subtype 1a infection in West Africa and the exponential phase represents introduction into new transmission networks in the developed world (see Fig. 1). Interestingly, the transition to exponential growth coincides with the appearance of three clades in the 1a phylogeny (Figs. 2a and 3a), suggesting three separate seeding or founder events. We estimate that exponential growth began between 1924 and 1953, and one could speculate that this was initiated by introduction of virus into opiate-injecting communities, such as that documented in London during the 1950s (Robson, 1999). The epidemic curve previously estimated from global subtype 1a infections did not resolve into two phases (Pybus et al., 2001). This difference may correctly represent the different populations sampled (i.e. the global HCV subtype 1a infected-population, as compared to the IDU risk group). Alternatively, it may reflect the increased length of NS5B sequences used in this study, which increases the power of the methods used.

The estimate for subtype 3a indicates that there has been a period of epidemic growth in this subtype starting around the first few decades of the 20th century. The transmission network responsible for the population growth at this time is unknown. It could have involved medical interventions or population movements within Asia (where HCV genotype 3 is thought to have originated) (Simmonds et al., 1996; Shah et al., 1997), or transmission of the virus into new limited populations in industrialised nations, with subsequent transmission into the growing IDU risk group. Injecting drug use is well described amongst soldiers in the First World War and it is possible that this could have contributed to early HCV transmission (Robson, 1999).

The recent decrease in growth rate observed for subtype 3a was not statistically significant and therefore should be viewed with caution. However, if this reduction is real and subtype-specific then the relative proportion of new infections will shift towards subtype 1a. Given the resistance of subtype 1a to currently available treatment (McHutchison et al., 1998), this would have the effect of increasing the disease burden per infection. Possible causes of the decreased growth rate for subtype 3a include risk-group saturation, clearance of HCV through treatment (which would be expected to favour subtype 3a), and reduction in transmission through behavioural policies such as needle exchange (Goldberg et al., 2001).

The growth rate of the exponential phase of the epidemic (r) is similar for subtypes 1a and 3a, representing a doubling of the number of HCV infections approximately every seven years. The similarity in growth rate between the two subtypes argues against any major subtype-specific differences in infectiousness within the IDU risk group. The “ R_0 values” presented are estimated from r using plausible values for the duration of infectiousness. When compared to published values of R_0 for other viruses, our estimates for HCV are similar to estimates for HIV transmission among male homosexuals in the UK (R_0 2–5), but less than estimates for HIV transmission between heterosexuals in Kampala (R_0 10–11) (Anderson, 1991). We could find no previous published estimates of R_0 for the transmission of HIV or other viruses within the IDU population.

Empirical observations of HCV spread amongst IDUs suggest that our estimated R_0 values do not represent the “basic reproductive number” of the virus, i.e. its intrinsic transmission potential. New IDUs acquire the virus very

rapidly after the initiation of needle-sharing (Bell et al., 1990; Garfein et al., 1996) and the seroprevalence of HCV in many IDU populations reaches 80–90% (Garfein et al., 1996; Goldberg et al., 1998). These findings indicate that R_0 in the IDU risk group is significantly greater than 2–3 and predict that in a static population equilibrium prevalence will be reached very rapidly (mathematically, the “effective reproductive number”, R_e , is expected to be close to 1 over a timescale of decades). Taken together with our observation of long-term epidemic growth, it appears that HCV transmission has been constrained by IDU population size since the beginning of the epidemic, and that there has been ongoing growth of the IDU population. In other words, the growth of HCV-infected IDUs has followed the growth of the IDU population. A long-term continual increase in the number of infections will always occur if the rate at which individuals move into the risk group and become infected is greater than the rate at which infected individuals cease to be infected (whether still in the risk group or not). Thus, our “ R_0 values” represent long-term growth rate of new HCV IDU infections, but underestimate the basic reproductive rate that would be achieved if the virus entered a new HCV-naïve IDU population.

The above interpretation of the HCV IDU epidemic predicts that the growth rate for HCV-infections obtained here should be similar to the growth rate of the IDU population. This is supported by epidemiological data from the UK. IDU prevalence in Glasgow, Scotland, has been estimated every five years since 1960 using a combination of consensus expert opinion and empirical data (pers. comm. Sharon Hutchinson). We fitted a model of exponential increase to these estimated numbers of IDUs using linear regression, which provided a good fit (coefficient of determination $R^2 = 0.9318$; $p < 0.01$). This longitudinal prevalence data indicated an exponential growth rate of 0.111 (0.1–0.14), corresponding to a doubling of the number of IDUs every 5–7 years. This rate is very similar to the growth rate of HCV infections estimated here using genetic data (Table 2).

The timing of the epidemic histories and the values of r presented were calculated using a nucleotide substitution rate estimated from an HCV subtype 1b outbreak (Power et al., 1994). This outbreak was exceptional in three ways: (i) it enabled us to observe evolution independently in multiple patients, (ii) the period of evolution observed was long (17 years), and (iii) the virus causing the outbreak was almost homogeneous. All these factors increase the accuracy of the evolutionary rate estimate, and the subtype 1b rate remains the preferred estimate until similar cohorts for subtypes 1a and 3a become available. The subtype 1b rate can clearly be generalised to other HCV genotypes, since it has been used to correctly estimate known events in the epidemic history of genotype 4 (Pybus et al., 2003).

As indicated in the Section 2, the coalescent framework assumes that: (i) the sample size is considerably smaller than the effective size of the population studied, (ii) genetic

distance is proportional to time, and (iii) the shape of the genealogy is unaffected by recombination, selection and population subdivision. The first of these assumptions is clearly upheld in the current study, the latter two are discussed more fully below.

Direct proportionality between genetic distance and time equates to a constant rate of nucleotide substitution. This assumption was tested using a likelihood ratio test and was accepted for subtype 3a but not subtype 1a. Previous analysis of HCV using phylogenies estimated from NS5B and E1 genes found that the molecular clock was rejected in about half the datasets studied (Pybus et al., 2001). In most cases, the clock was rejected for one gene but not the other, yet the epidemic histories predicted by the two genes were not significantly different, indicating that the degree of rate variation present was not sufficient to bias results. A recent comprehensive analysis of nucleotide rate variation in RNA viruses also shows that substitution rates estimated using the maximum likelihood method are reliable indicators of the evolutionary timescale when small amounts of rate heterogeneity are present (Jenkins et al., 2002).

Regarding influences on the shape of the HCV genealogy, recombination has been shown to occur in HCV but appears to be rare (Kalinina et al., 2002). We cannot formally exclude selection influences, but previous genealogies for HCV subtype 1a constructed from NS5b sequences did not differ significantly from genealogies constructed from E1 sequences, although these two genes are under different selection pressures (Pybus et al., 2001). Given that HCV in the IDU population is transmitted by needle sharing one might expect the genealogy to be structured by subdivisions based on city or region of origin. However, within the UK, extensive phylogenetic mixing between HCV sequences amplified from geographically dispersed IDU communities has been demonstrated (Cochrane et al., 2002). Also, HCV infections do not appear to be structured by host age (Cochrane et al., 2002), although structuring by other social groupings cannot be wholly excluded. There is less phylogenetic mixing between sequences from the UK and those from France, and significant segregation is observed between European and Australian sequences (Cochrane et al., 2002). Given this segregation, analysis of the extended dataset does violate the assumption of no population subdivision. However, the addition of the Marseilles and Melbourne sequences did not significantly alter the results. The HCV epidemic in these locations could be analysed separately once more data becomes available (>25 sequences is preferable).

In conclusion, a new method was employed to investigate the history of the HCV subtype 1a and 3a epidemic in IDUs. For HCV subtype 1a, the history comprises an endemic phase followed by exponential growth, and suggests that the epidemic in the UK was seeded by at least three separate introductions from an ancestral population. Both subtypes demonstrate a period of exponential growth in the 20th century with a doubling time of approximately seven to eight years. We show that the rate of spread is likely to have been

constrained by the size of the IDU population, such that our estimates reflect the growth of the IDU risk group rather than the innate transmission potential of the virus.

Acknowledgements

Thanks to S. Bowden, (Victoria Infectious Diseases Reference Laboratory, Melbourne), S. Sutherland, G. Leadbetter (Regional Virus Laboratory, Edinburgh), L.

MacDonald, G. Cunningham (Regional Virus Laboratory, Glasgow), R. Jacobs (Royal Free Hospital, London) and P. Grant (University College, London) for supplying samples, and S. Hutchinson (Scottish Centre for Infection and Environmental Health) for data regarding IDU in Glasgow. This work was funded by Scottish Hospitals Endowment Research Trust (Grant RG1/00) and The Wellcome Trust (Grant 66889). Thanks to Nick Grassly for very helpful critical reading.

Appendix A

Details of the demographic models used in the coalescent analysis are provided here. $N(t)$ denotes the effective number of infections at time t (t increases into the past, hence $t = 0$ is the time at which the sequences were sampled).

Exponential growth

Equation:
$$N(t) = N(0) \exp[-rt]$$

Parameters: $N(0)$ = effective number of infections at $t = 0$
 r = exponential growth rate

Logistic growth

Equation:
$$N(t) = N(0) \frac{1 + c}{1 + c \exp[rt]}$$

Parameters: $N(0)$ = effective number of infections at $t = 0$
 r = exponential growth rate when $N(t)$ is small
 c = shape parameter, determining the level of density-dependence at $t = 0$

Piecewise expansion growth

Equation:
$$N(t) = \begin{cases} N(0)\exp[-rt] & \text{if } t < -\ln(\alpha)/r \\ N(0)\alpha & \text{otherwise} \end{cases}$$

Parameters: $N(0)$ = effective number of infections at $t = 0$
 r = exponential growth rate
 α = population size prior to exponential growth, as a proportion of $N(0)$

References

- Alter, M.J., Hutin, Y.J., Armstrong, G.L., 2001. Epidemiology. In: Liang, T., Hoofnagle, J. (Eds.), *Hepatitis C*. Academic Press, pp. 169–183.
- Anderson, R.M., 1991. Transmission dynamics and control of infectious disease agents. In: Anderson, R.M., May, R.M. (Eds.), *Population Biology of Infectious Diseases*, Springer, Berlin, pp. 149–176.
- Anderson, R.M., May, R.M., 1991. A framework for the population biology of infectious diseases. In: Anderson, R.M., May, R.M. (Eds.), *Infectious Diseases of Humans: Dynamics and Control*, Springer, Berlin.
- Bell, J., Batey, R.G., Farrell, G.C., Crewe, E.B., Cunningham, A.L., Byth, K., 1990. Hepatitis C virus in intravenous drug users. *Med. J. Aust.* 153, 274–276.
- Cochrane, A., Searle, B., Hardie, A., Robertson, R., Delahooke, T., Cameron, S., et al., 2002. A genetic analysis of hepatitis C virus transmission between injection drug users. *J. Infect. Dis.* 186, 1212–1221.
- Crofts, N., Aitken, C.K., Kaldor, J.M., 1999. The force of numbers: why hepatitis C is spreading among Australian injecting drug users while HIV is not. *Med. J. Aust.* 170, 220–221.
- Donnelly, P., Tavaré, S., 1995. Coalescents and genealogical structure under neutrality. *Annu. Rev. Genet.* 29, 401–421.
- Felsenstein, J., 1981. Evolutionary trees from DNA sequences: a maximum likelihood approach. *J. Mol. Evol.* 17, 368–376.
- Garfein, R.S., Vlahov, D., Galai, N., Doherty, M.C., Nelson, K.E., 1996. Viral infections in short-term injection drug users: the prevalence of the hepatitis C, hepatitis B, human immunodeficiency, and human T-lymphotropic viruses. *Am. J. Public Health* 86, 655–661.
- Goldberg, D., Cameron, S., McMenamin, J., 1998. Hepatitis C antibody prevalence among injecting drug users in Glasgow has fallen but remains high. *Commun. Dis. Public Health* 1, 95–97.
- Goldberg, D., Burns, S., Taylor, A., Cameron, S., Hargreaves, D., Hutchinson, S., 2001. Trends in HCV prevalence among injecting drug users in Glasgow and Edinburgh during the era of needle exchange. *Scand. J. Infect. Dis.* 33, 457–461.
- Holmes, E.C., Nee, S., Rambaut, A., Garnett, G.P., Harvey, P.H., 1995. Revealing the history of infectious disease epidemics using phylogenetic trees. *Philos. Trans. R. Soc. London B* 349, 33–40.
- Jeannel, D., Fretz, C., Traore, Y., Kohdjo, N., Bigot, A., Gamy, E.P., et al., 1998. Evidence for high genetic diversity and long-term endemicity of

- hepatitis C virus genotypes 1 and 2 in West Africa. *J. Med. Virol.* 55, 92–97.
- Jenkins, G.M., Rambaut, A., Pybus, O.G., Holmes, E.C., 2002. Rates of molecular evolution in RNA viruses: a quantitative phylogenetic analysis. *J. Mol. Evol.* 54, 156–165.
- Joy, D.A., Feng, X., Mu, J., Furuya, T., Chotivanich, K., Krettli, A.U., Ho, M., Wang, A., White, N.J., Suh, E., Beerli, P., Su, X.-Z., 2003. Early origin and recent expansion of *Plasmodium falciparum*. *Science* 300, 318–321.
- Kalinina, O., Norder, H., Mukomolov, S., Magnius, L.O., 2002. A natural intergenotypic recombinant of hepatitis C virus identified in St. Petersburg. *J. Virol.* 76, 4034–4043.
- Leigh, J.P., Bowler, C.L., Leistikow, B.N., Schenker, M., 2001. Costs of hepatitis C. *Arch. Intern. Med.* 161, 2231–2237.
- Lemey, P., Pybus, O.G., Wang, B., Saksena, N.K., Salemi, M., Vandamme, A.M., 2003. Tracing the origin and history of the HIV-2 epidemic. *Proc. Natl. Acad. Sci. U.S.A.* 100, 6588–6659.
- McHutchison, J.G., Gordon, S.C., Schiff, E.R., Shiffman, M.L., Lee, W.M., Rustgi, V.K., et al., 1998. Interferon alfa-2b alone or in combination with ribavirin as initial treatment for chronic hepatitis C. *N. Engl. J. Med.* 339, 1485–1492.
- Nakano, T., Lu, L., Liu, P., Pybus, O.G., 2004. Viral gene sequences reveal the variable history of hepatitis C virus infection among countries. *J. Infect. Dis.* 190, 1098–1108.
- Power, J.P., Lawlor, E., Davidson, F., Holmes, E.C., Yap, P.L., Simmonds, P., 1995. Molecular epidemiology of an outbreak of infection with hepatitis C virus in recipients of anti-D immunoglobulin. *Lancet.* 345, 1211–1213.
- Pybus, O.G., Rambaut, A., Harvey, P.H., 2000. An integrated framework for the inference of viral population history from reconstructed genealogies. *Genetics* 155, 1429–1437.
- Pybus, O.G., Charleston, M.A., Gupta, S., Rambaut, A., Holmes, E.C., Harvey, P.H., 2001. The epidemic behaviour of hepatitis C virus. *Science* 22, 2323–2325.
- Pybus, O.G., Rambaut, A., 2002. GENIE: estimating demographic history from molecular phylogenies. *Bioinformatics* 18, 1404–1405.
- Pybus, O.G., Drummond, A.J., Nakano, T., Robertson, B.H., Rambaut, A., 2003. The epidemiology and iatrogenic transmission of hepatitis C virus in Egypt: a bayesian coalescent approach. *Mol. Biol. Evol.* 20, 381–387.
- Rambaut, A., 2000. Estimating the rate of molecular evolution: incorporating non-contemporaneous sequences into maximum likelihood phylogenies. *Bioinformatics* 16, 395–399.
- Robson, P., 1999. Heroin and the opioids, historical background, *Forbidden Drugs*, second ed. Oxford University Press, Oxford.
- Schreiber, G.B., Busch, M.P., Kleinman, S.H., Korelitz, J.J., 1996. The risk of transfusion-transmitted viral infections. *N. Engl. J. Med.* 334, 1685–1690.
- Shah, H.A., Jafri, W., Malik, I., Prescott, L., Simmonds, P., 1997. Hepatitis C virus (HCV) genotypes and chronic liver disease in Pakistan. *J. Gastroenterol. Hepatol.* 12, 758–761.
- Simmonds, P., Mellor, J., Sakuldamrongpanich, T., Nuchaprayoon, C., Tanprasert, S., Holmes, E.C., et al., 1996. Evolutionary analysis of variants of hepatitis C virus found in South-East Asia: comparison with classifications based upon sequence similarity. *J. Gen. Virol.* 77, 3013–3024.
- Smith, D.B., Pathirana, S., Davidson, F., Lawlor, E., Power, J., Yap, P.L., et al., 1997. The origin of hepatitis C virus genotypes. *J. Gen. Virol.* 78, 321–328.
- Strimmer, K., Pybus, O.G., 2001. Exploring the demographic history of DNA sequences using the generalized skyline plot. *Mol. Biol. Evol.* 18, 2298–2305.
- Swofford, D.L., 1998. PAUP*. *Phylogenetic Analysis Using Parsimony (*and Other Methods)*. Version 4.
- Tanaka, Y., Hanada, K., Mizokami, M., Yeo, A.E.T., Shih, J., Gojobori, T., Alter, H.J., 2002. 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.* 99, 15584–15589.
- Tanaka, Y., Agha, S., Saady, N., Kurbanov, F., Orito, E., Kato, T., Abo-Zeid, M., Khalaf, M., Miyakawa, Y., Mizokami, M., 2004. Exponential spread of hepatitis C virus genotype 4a in Egypt. *J. Mol. Evol.* 58, 191–195.
- Taylor, A., Goldberg, D., Hutchison, S., Cameron, S., Gore, S., McMena-min, J., et al., 2000. Prevalence of hepatitis C virus infection amongst injecting drug users in Glasgow 1990–1996: are current harm reduction strategies working? *J. Infect.* 40, 176–183.
- Trepo, C., Pradat, P., 1999. Hepatitis C virus infection in Western Europe. *J. Hepatol.* 31 (Suppl. 1), 80–83.