

# The Epidemiology and Iatrogenic Transmission of Hepatitis C Virus in Egypt: A Bayesian Coalescent Approach

O. G. Pybus,\* A. J. Drummond,† T. Nakano,‡ B. H. Robertson,‡ and A. Rambaut\*

\*Department of Zoology, University of Oxford, Oxford, United Kingdom; †School of Biological Sciences, University of Auckland, Auckland, New Zealand; and ‡Division of Viral Hepatitis, Centers for Disease Control and Prevention, Atlanta, Georgia

Hepatitis C virus (HCV) is a leading cause of liver cancer and cirrhosis, and Egypt has possibly the highest HCV prevalence worldwide. In this article we use a newly developed Bayesian inference framework to estimate the transmission dynamics of HCV in Egypt from sampled viral gene sequences, and to predict the public health impact of the virus. Our results indicate that the effective number of HCV infections in Egypt underwent rapid exponential growth between 1930 and 1955. The timing and speed of this spread provides quantitative genetic evidence that the Egyptian HCV epidemic was initiated and propagated by extensive antischistosomiasis injection campaigns. Although our results show that HCV transmission has since decreased, we conclude that HCV is likely to remain prevalent in Egypt for several decades. Our combined population genetic and epidemiological analysis provides detailed estimates of historical changes in Egyptian HCV prevalence. Because our results are consistent with a demographic scenario specified a priori, they also provide an objective test of inference methods based on the coalescent process.

## Introduction

Hepatitis C Virus (HCV) is recognized as a major threat to global public health. An estimated 170 million people worldwide are infected, most of them chronically infected and at risk for liver cirrhosis and hepatocellular carcinoma (HCC) (World Health Organization 1997a). HCV-associated mortality is expected to increase substantially in the next 20 years (Centers for Disease Control and Prevention 1998). To be successful, HCV treatment, prevention, and vaccination strategies must be based on a sound understanding of the evolutionary and epidemiological behavior of the virus, especially in those geographic regions most seriously affected.

Egypt has possibly the highest HCV prevalence in the world; 10%–20% of the general population are infected and HCV is the leading cause of HCC and chronic liver disease in the country (Arthur et al. 1997; Nafeh et al. 2000; El-Zayadi et al. 2001; Habib et al. 2001; Hassan et al. 2001). Approximately 90% of Egyptian HCV isolates belong to a single subtype, 4a, which responds less successfully to interferon therapy than other subtypes (El-Zayadi et al. 1996; Angelico et al. 1997; Ray et al. 2000; Kamal et al. 2000). Furthermore, HCV is less prevalent in countries neighboring Egypt that have similar socio-medical conditions and similar HCV strains (McCarthy et al. 1994; World Health Organization 1997b; Shobokshi et al. 1999).

Why, then, is Egypt so seriously affected? Previous research has suggested that the Egyptian HCV epidemic results from the use of unsterile injection equipment during mass treatment of the general population with parenteral antischistosomal therapy (PAT) (Quinti et al. 1995; Frank et al. 2000). PAT was extensively practiced in Egypt from the 1920s to the 1980s and was gradually replaced by oral treatment from the 1970s onward. The most common PAT drug, tartar emetic (potassium antimony tartarate), was administered over the course of a few weeks as a series of intravenous injections (Maegraith 1964). The potential for

transmission of blood-borne pathogens in such circumstances is considerable. Cross-sectional epidemiological analyses have provided evidence for the PAT hypothesis; there is a correlation between the level of exposure to PAT and HCV prevalence among different age groups and geographic regions (Angelico et al. 1997; Frank et al. 2000; Nafeh et al. 2000; Habib et al. 2001).

Although subtype 4a is the dominant Egyptian HCV strain, a survey by Ray et al. (2000) of HCV genetic diversity in the country revealed that other subtypes (provisionally named 4 $\alpha$ , 4 $\beta$ , and 1g) are also present at lower prevalences. The most recent common ancestor of each subtype (including 4a) existed approximately 80–120 years ago, suggesting the recent and simultaneous appearance of a few pre-existing endemic strains. This is in contrast to the pattern of HCV genetic diversity found in other developing regions, which is more consistent with a long period of endemic infection (Smith et al. 1997; Pybus et al. 2001).

Standard analyses of genetic diversity, however, use only a fraction of the information contained in viral gene sequences. We have previously developed a framework based on coalescent theory that reconstructs the population dynamic history of HCV from sampled sequences, enabling us to date epidemiological events and estimate historical rates of epidemic spread (Pybus et al. 2000, 2001). Here we adapt this framework specifically to investigate the epidemic history of HCV in Egypt. Our results are obtained using a Bayesian inference framework that implements a detailed nucleotide-substitution model and incorporates the uncertainty inherent in phylogenetic reconstruction (Drummond et al. 2002). This method represents a significant improvement on previous analyses that made inferences from a single phylogeny and therefore ignored phylogenetic error (Pybus et al. 2000).

Our results reveal a history of epidemic spread that is unique to Egypt, and they provide strong genetic evidence for the role of PAT in the initiation and propagation of HCV in Egypt. In addition, the quantitative corroboration of our results with a specific epidemiological scenario provides an independent test of inference methods based on the coalescent process.

Key words: hepatitis C, Egypt, epidemiology, Bayesian inference.

E-mail: oliver.pybus@zoo.ox.ac.uk.

*Mol. Biol. Evol.* 20(3):381–387. 2003

DOI: 10.1093/molbev/msg043

© 2003 by the Society for Molecular Biology and Evolution. ISSN: 0737-4038

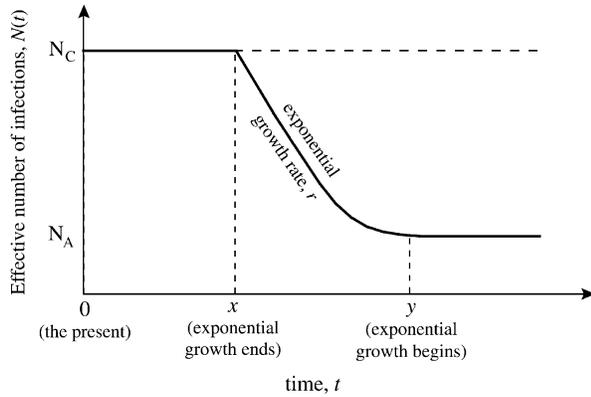


FIG. 1.—The demographic model (equation 2). The model has five demographic parameters,  $N_C$ ,  $N_A$ ,  $x$ ,  $y$ , and  $r$ , only four of which are needed to fully specify the model. The effective number of infections is a compound variable that is considered to be linearly proportional to the true number of infections.

## Methods

### General Statistical Framework

The past population dynamics of a virus can be inferred from viral gene sequence data using a population genetic model called *the coalescent* (Kingman 1982; Griffiths and Tavaré 1994). The coalescent framework requires a demographic model, denoted  $N(t)$ , that describes the effective population size through time. Demographic models can represent either increasing or decreasing populations; commonly used forms are constant size, exponential growth, population bottlenecks, and logistic growth. The models contain one or more demographic parameters, collectively denoted by  $\Theta$ . Time ( $t$ ) is defined in reverse, so that the most recent sequence was sampled at time zero and  $t$  increases into the past. Given a set of sampled gene sequences, it is possible to estimate the parameters  $\Theta$  using a variety of statistical methodologies (e.g., Griffiths and Tavaré 1994; Kuhner, Yamato, and Felsenstein 1998; Grassly et al. 1999; Pybus, Rambaut, and Harvey 2000; Drummond et al. 2002).

In this article we use sample-based Bayesian inference. Drummond et al. (2002) recently described a Metropolis-Hastings Markov chain Monte Carlo (MCMC) method for the joint estimation of genealogy ( $g$ ), demographic parameters ( $\Theta$ ), and mutational parameters ( $\Phi$ ). This approach is implemented in the computer program MEPI (Version 1.0) available from <http://cebl.auckland.ac.nz/mepi>. Given the sampled gene sequences  $D$ , parameter estimates are calculated by averaging over many genealogies and weighting the contribution of each genealogy by its likelihood given the data. The MCMC algorithm performs this calculation efficiently by sampling the posterior probability density  $h_{\Theta\Phi g}$ :

$$\begin{aligned} h_{\Theta\Phi g}(\Theta, \Phi, g \mid D, \mu) \\ = \frac{1}{Z} Pr\{D \mid \Phi, g\} \cdot f(g \mid \Theta) \cdot p(\Theta) \cdot q(\Phi). \end{aligned} \quad (1)$$

$Pr\{D \mid \Phi, g\}$  is the likelihood of genealogy  $g$  given a specified nucleotide substitution model, the parameters of which are collectively denoted by  $\Phi$  (Felsenstein 1981).

MEPI implements the Jukes-Cantor, HKY and GTR substitution models, as well as the gamma and codon site-specific models of rate heterogeneity among sites. In addition, the genealogical likelihood  $Pr\{D \mid \Phi, g\}$  is extended in MEPI in order to incorporate the sampling time of noncontemporary sequences (Rambaut 2000). This enables the mutation rate and effective population size parameters to be estimated separately, although this option was not employed in our analysis because the sequences used were all sampled at the same time.  $f(g \mid \Theta)$  is the probability density of the coalescent model,  $p(\Theta)$  is the prior distribution of the demographic parameters  $\Theta$ ,  $q(\Phi)$  is the prior distribution of the mutational parameters  $\Phi$ , and  $Z$  is the normalizing constant. MEPI does not currently incorporate models of population subdivision, although this is planned for future versions (Drummond et al. 2002).

In the analysis presented below, we are interested in the marginal probability density  $h_{\Theta\Phi}(g \mid D, \mu)$  obtained by sampling from the full posterior probability density  $h_{\Theta\Phi g}$ . Thus the genealogical information is statistically regarded as “missing data.” The following sections provide details of the specific demographic and mutational models used in our analysis of HCV in Egypt.

### Demographic Model

Pybus et al. (2001) previously presented a simple demographic model for HCV epidemic behavior. This model assumed a continuous epidemic process in which the viral transmission parameters (such as probability of transmission per contact event) remain constant through time. Such a model is obviously not suitable for HCV in Egypt, which appears to have followed a discontinuous process, putatively as a result of past PAT campaigns. We have therefore developed a new demographic model specifically to investigate the Egyptian HCV epidemic.

$$N(t) = \begin{cases} N_C & \text{if } t \leq x \\ N_C e^{-r(t-x)} & \text{if } x < t < y \\ N_A & \text{if } t \geq y \end{cases}. \quad (2)$$

This model, illustrated in figure 1, allows an ancestral population to undergo exponential growth over a finite time period, resulting in a larger modern population.  $N_C$  is the current effective number of infections, which is constant for all times later than time  $x$ .  $N_A$  is the ancestral effective number of infections, which is constant for all times earlier than time  $y$ . Between times  $x$  and  $y$  the population grows exponentially at rate  $r$ . Although five parameters are given, only four are needed to completely specify the model. For example, the ancestral population size  $N_A$  can be calculated as  $N_A = N_C e^{-r(y-x)}$ . We confirmed that equation 2 provided a good fit to the demographic signal in the data by applying a coalescent model-selection algorithm, the generalized skyline plot (Strimmer and Pybus 2001).

### Mutational Model

As the Egyptian HCV gene sequences used were all sampled at the same time (see below) the mean mutation rate ( $\mu$ ) and population size ( $N_C$  and  $N_A$ ) parameters could

only be separated if prior information about one of these parameters was available. Therefore our estimates of epidemic history were rescaled into units of calendar years by fixing the mean mutation rate  $\mu$  to a previously estimated value of  $0.79 \times 10^{-3}$  substitutions/site/year (Pybus et al. 2001). This rate was estimated for the E1 gene of HCV, which is the genome region used in this study.

To select an appropriate nucleotide-substitution model for our data, we used an estimated maximum likelihood tree to compare the genealogical likelihood  $\Pr\{D|\Phi, g\}$  under various substitution models, in a manner similar to that employed by the Modeltest program (Posada and Crandall 1998). Allowing for rate heterogeneity among sites massively improved the log likelihood, and also increased the time to the most recent common ancestor (TMRCA) of the sequences. Thus, in our full Bayesian analysis, we implemented a codon site-specific model of rate heterogeneity. A much smaller improvement in likelihood was obtained when the HKY and GTR substitution models were compared, although this difference was statistically significant according to a likelihood ratio test. As a result, the full MCMC analysis was performed twice, once using HKY and once using GTR. The results were qualitatively identical and quantitatively very similar, so we report here only the results obtained using the HKY model (the GTR results are available on request).

Although the mean mutation rate was fixed, we were able to estimate absolute mutation rates for each codon position (denoted  $\mu_1$ ,  $\mu_2$  and  $\mu_3$ ) using MCMC integration. This was achieved using two proposal mechanisms (a random walk and a centered linear transformation) that sampled the mutational parameter space fully and efficiently while maintaining the constraint  $(\mu_1 + \mu_2 + \mu_3)/3 = \mu$ . We also estimated the HKY and GTR model parameters and the TMRCA of the sequences. MCMC convergence was tested as in Drummond et al. (2002).

#### Prior Distributions

A uniform prior distribution on  $r$  was used, with a lower bound of zero and an upper bound of 0.75. The upper bound represents a maximum plausible growth rate, corresponding to an epidemic doubling time of less than one year. As discussed in *Results*, this bound was impinged on in the MCMC analysis. However, the posterior probability density of  $r$  was focused on substantially lower rates and the lower bound was not reached (fig. 2). All other prior distributions were uniform and no other parameter boundaries were impinged upon.

#### Sequence Data

Two HCV data sets of partial E1 gene sequences (including 42 nucleotides from the neighboring Core gene) were obtained from a comprehensive study of HCV diversity in Egypt (Ray et al. 2000). These sequences are highly suitable for coalescent analysis for several reasons: (i) they represent a geographically diverse and approximately random sample from the study population, (ii) they

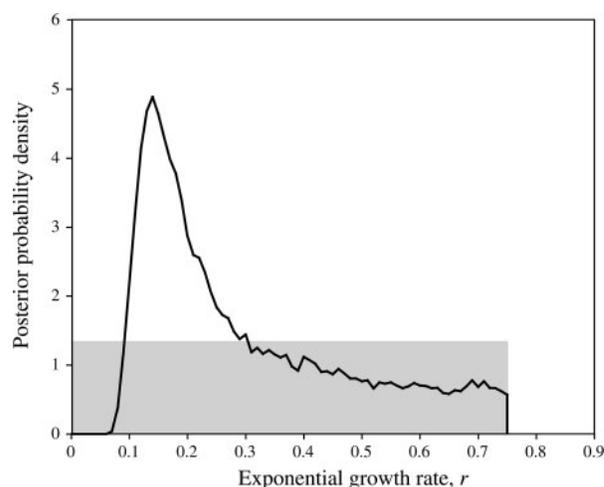


FIG. 2.—The estimated distribution of the exponential growth rate,  $r$  (black line), calculated using data set B. The distribution shows that values of  $r$  smaller than 0.07 are not supported by the data. Very similar results were obtained for data set A. The shaded area represents the uniform prior distribution of  $r$  that was used.

show no correlation between genetic and geographic distance (i.e., no obvious population subdivision; Ray et al. 2000), (iii) the sequences contain ample phylogenetic information, (iv) the sampling date (1993) is known, and (v) an independent estimate of nucleotide-substitution rate for the gene region used is available (see above).

The first data set, labeled A, contains 68 sequences 411bp in length (63 type 4 sequences and 5 subtype 1g sequences). This represents all the E1 sequences published by Ray et al. (2000) except for three isolates belonging to subtypes 1a or 1b, which are recent introductions into Egypt rather than endemic strains. Although HCV type 4 is the dominant type in Egypt and appears to have been present in the Middle East for several centuries (Smith et al. 1997), it is not clear whether subtype 1g is an endemic or introduced strain. Thus the shared history of type 4 and subtype 1g could possibly represent a non-Egyptian ancestral population. To exclude any potential estimation bias arising from this, we created a second data set, labeled B, which contains the 63 type 4 sequences only.

#### Results

The estimated demographic and mutational parameter values for data sets A and B are shown in table 1. With the exception of the date of the most recent common ancestor, the two data sets give very similar results. This indicates that our demographic parameters are robust to the inclusion or exclusion of the highly divergent subtype 1g strains (see above). Data set A includes these strains and thus has a significantly older common ancestor than data set B, which contains type 4 sequences only. Our results suggest that the most recent ancestor of HCV type 4 in Egypt existed around 300 years ago.

The demographic parameter estimates indicate that HCV infection in Egypt changed from a low endemic level to rapid exponential growth sometime during the early 1930s. This period of exponential growth continued until

**Table 1**  
**Parameter Estimates for the Egyptian HCV Data Sets**

Parameter	Estimated Parameter Values <sup>a</sup> (95% confidence region) <sup>b</sup>	
	Data Set A	Data Set B
Demographic parameters		
$r$ (exponential growth rate)	0.264 (0.075, 0.620)	0.237 (0.072, 0.564)
$y$ (exponential growth start date)	1934 (1924, 1943)	1932 (1922, 1940)
$x$ (exponential growth end date)	1953 (1941, 1966)	1953 (1941, 1966)
$N_C$ (current effective number of infections)	10310 (4095, 18960)	8779 (3323, 15780)
$N_A$ (ancestral effective number of infections)	245 (153, 345)	170 (99.6, 251)
Mutational parameters		
Mutation rates (substitutions per site per year)		
$\mu_1$ (codon position 1)	0.45e-3 (0.40e-3, 0.49e-3)	0.45e-3 (0.39e-3, 0.49e-3)
$\mu_2$ (codon position 2)	0.23e-3 (0.198e-3, 0.27e-3)	0.25e-3 (0.21e-3, 0.28e-3)
$\mu_3$ (codon position 3)	1.69e-3 (1.63e-3, 1.75e-3)	1.68e-3 (1.62e-3, 1.74e-3)
$\kappa$ (transition/transversion rate ratio)	7.71 (6.75, 8.69)	8.26 (7.13, 9.38)
Date of most recent common ancestor	1374 (1258, 1481)	1710 (1673, 1747)

<sup>a</sup> Parameter estimates are mean posterior density values.

<sup>b</sup> Confidence regions are highest posterior density intervals.

the 1950s and led to an approximate 50-fold increase in the effective number of infections.

The mean posterior estimate of the HCV growth rate in Egypt is  $r = 0.25 \text{ year}^{-1}$ . This point estimate is not particularly informative because the posterior distribution of  $r$  is right skewed (see fig. 2). Because the posterior

density is focused in the range 0.1–0.3, it would be more appropriate use this range as an estimate for  $r$ . Thus the spread of HCV in Egypt was rapid, with an estimated epidemic doubling time of 2 to 7 years. Even the lower confidence limit of the Egyptian growth rate ( $r \sim 0.07$ ) is greater than the rate estimated for type 4 infections outside Egypt ( $r \sim 0.04$ ) (Pybus et al. 2001). In fact, the growth rate in Egypt is as fast, or faster, than that estimated for HCV subtypes 1a and 1b ( $r \sim 0.09$ ), which have spread primarily through highly efficient parenteral transmission networks, namely injecting drug use and infected blood products (Pybus et al. 2001). Thus our growth rate result indicates that an equally efficient route must have been at work during the twentieth century in Egypt. Parenteral transmission via PAT treatment is the only viable candidate route.

As shown in figure 2, the upper bound on the prior distribution for  $r$  was impinged upon in the MCMC analysis, and the posterior distribution of  $r$  is right skewed, indicating that the data do not reject exceptionally high growth rates. However, the lower bound of zero was not reached, and we can further conclude that  $r$  is at least 0.07. Thus the data do strongly support a switch from low to high numbers of infections sometime between 1924 and 1966, but do not contain much information about the upper limit of the epidemic growth rate.

The mutational parameter estimates show a strong bias toward transitions over transversions in the E1 gene region, as previously reported (Salemi and Vandamme 2002). Our estimates of codon-position-specific mutation rates indicate a faster rate at third codon positions, which is commonly observed and is most likely due to the increased opportunity for unconstrained “silent” (synonymous) nucleotide substitution at third positions.

By sampling parameter values from the marginal probability density  $h_{\Theta\Phi}(\Theta\Phi|D,\mu)$  we were able to reconstruct an estimated growth curve for the Egyptian HCV epidemic (fig. 3). The reconstructed epidemic histories are similar for both data sets and show a rapid increase from 1930 to 1960 in the effective number of infections. This growth period lies entirely within the period of widespread

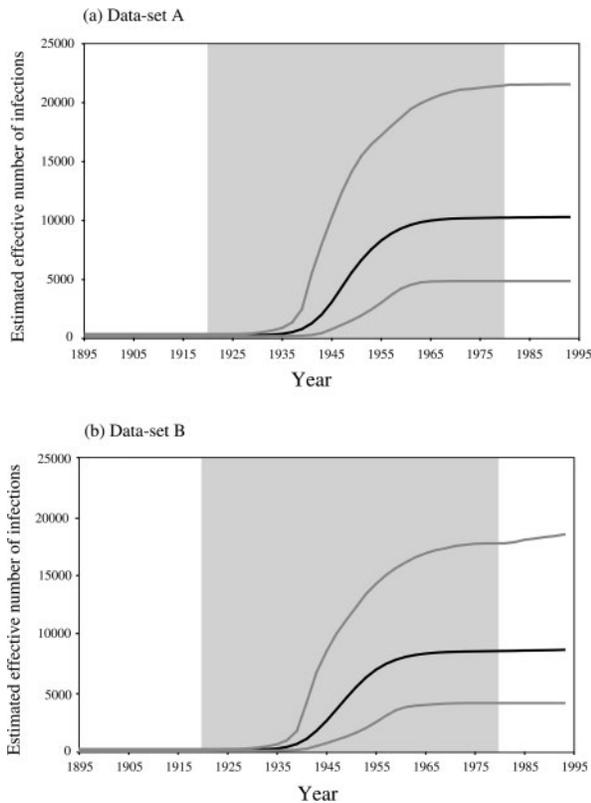


Fig. 3.—The reconstructed epidemic growth curve for the Egyptian HCV epidemic (black line), with estimated highest posterior density confidence intervals (gray lines). The shaded area indicates the period of widespread PAT treatment in Egypt. (a) Data set A. (b) Data set B. Obviously, these results are not independent as data set B is a subset of data set A.

PAT treatment in Egypt. In the context of cross-sectional epidemiological evidence (Frank et al. 2000), this temporal correlation constitutes independent genetic evidence for the involvement of PAT in the initiation and generation of the Egyptian HCV epidemic.

The estimated exponential growth period is shorter than the period of PAT treatment. This disparity could be due to a combination of epidemiological processes—the delay in exponential growth after initiation of PAT may result from stochastic or spatial effects that reduce mean growth rates (e.g., Lande 1998; May, Gupta, and McLean 2001), whereas the early end of exponential growth could have resulted from either saturation of the core risk group, or a decrease in the number of PAT patients from 1965 onward (Frank et al. 2000). Alternatively, our results may be limited by the amount of information contained in the single locus analyzed. In addition, levels of past PAT treatment and HCV prevalence vary considerably among different regions in Egypt, so our estimated growth rate represents an average rate for the country as a whole. We did not detect any significant difference in the epidemic history of sequences from Upper, Lower, and Middle Egypt (results not shown), although this is almost certainly due to the resulting small sample sizes.

## Conclusion

PAT has been assumed to be a major risk factor for HCV infection in Egypt (Quinti et al. 1995; Angelico et al. 1997; Frank et al. 2000). Our results support this hypothesis and indicate that the current high prevalence of HCV in Egypt is the result of rapid exponential epidemic growth that occurred during the period of PAT treatment. Furthermore, the speed of the epidemic points to an efficient, parenteral transmission route, and stands in contrast to the relatively ineffective social and domestic transmission routes that are thought to account for the endemic maintenance of HCV infection in other developing countries (Centers for Disease Control and Prevention 1998; Shobokshi et al. 1999; Kao and Chen 2000).

From an epidemiological perspective, the course of the Egyptian HCV epidemic resulted from a temporary increase in  $R_0$ , the basic reproductive number of the virus, during the PAT treatment period.  $R_0$  represents the average number of secondary infections created by one primary infection in a susceptible population. An increase in  $R_0$  can result from three changes: increased probability of transmission per contact event, more frequent exposure to infection, and a longer period of infectiousness. In the absence of adequate needle hygiene, it is clear that the injectable nature of PAT could produce the first change, while the widespread administration of PAT could produce the second. Prior to PAT, HCV infection in Egypt was probably at endemic equilibrium with an  $R_0$  close to 1.7, which is the value estimated for current HCV type 4 infections outside Egypt (Pybus et al. 2001). If the general relationship  $R_0 = rD + 1$  holds true for the exponential phase of the Egyptian epidemic, then  $R_0$  was raised to approximately 3–7 during the period of PAT treatment ( $D$  is the mean duration of infectiousness, and following Pybus et al. (2001) it is averaged across the range 10 to 30 years).

Hepatitis C virus infection in Egypt increases with age (e.g., Frank et al. 2000), suggesting that transmission has considerably decreased since the end of PAT treatment, such that the  $R_0$  of HCV in Egypt today has probably declined to its pre-PAT level. This temporal change in  $R_0$  provides further evidence that the epidemic behavior of HCV is primarily determined by the transmission routes and networks that individual HCV strains find themselves in. However, even if contemporary levels of transmissibility and exposure to infection are small, the high HCV prevalence in Egypt means that the absolute number of new infections per year may still be significant. The public health situation is compounded by the long duration of chronic HCV infection and frequent co-infection with schistosomes (Kamal et al. 2000; Hassan et al. 2001; Gad et al. 2001). Although our results show that infection via PAT occurred several decades ago, the slow development of HCV-related liver cirrhosis and cancer (Centers for Disease Control and Prevention 1998) means that the incidence of these diseases in Egypt may not yet have peaked. Even under the most optimistic scenario of zero contemporary transmission, a standard dynamical model of prevalence decay (Anderson and May 1991) indicates that HCV infection in Egypt will remain above 5% for at least the next 50 years. Transmission prevention strategies, vaccine development, and cheaper antiviral drugs are all required to reduce the current and future HCV disease burden in Egypt. Anti-viral combination drug therapy (interferon  $\alpha$ -2b or  $\alpha$ -2a plus ribavirin) is effective against HCV type 4 infections in 40%–60% of cases (El-Zayadi et al. 1999; Shobokshi et al. 2002), but the high cost of this treatment may affect the extent of its use in Egypt.

On a more theoretical note, the PAT hypothesis provides a quantitative a priori scenario and can therefore be used to judge the accuracy of parameter estimates obtained from sequence data using coalescent theory. Such test cases are relatively rare and enable us to objectively evaluate the performance of analytical methods. Coalescent theory has been similarly investigated in the past by Rodrigo et al. (1999), who used a coalescent model to estimate the generation time of HIV-1 (human immunodeficiency virus type 1) from viral gene sequence data. Their estimated generation time was close to that obtained using alternative nongenetic mathematical models. The correspondence of our estimated dates and growth rates with independently obtained epidemiological results also suggests that the coalescent framework is reliable in this case. Thus it appears that possible confounding factors (substitution rate variation among lineages, or epidemic differences among regions in Egypt) have not significantly biased our estimates. Viral recombination, which could potentially strongly affect coalescent analyses (Schierup and Hein 2000; Worobey 2001), is very rare in HCV, having only been detected once (Kalinina et al. 2002).

## Acknowledgments

This work was funded by the Wellcome Trust (O.G.P.), The Royal Society (A.R.), and The New Zealand Foundation for Research, Science and Technology

(A.J.D.). Many thanks to Paul Harvey, Eddie Holmes, Sunetra Gupta, the reviewers, and the associate editor for their helpful comments on the manuscript.

## Literature Cited

- Anderson, R. M., and R. M. May. 1991. *Infectious diseases of humans: dynamics and control*. Oxford University Press.
- Angelico, M., E. Renganathan, C. G. Gandin et al. (14 co-authors). 1997. Chronic liver disease in the Alexandria governorate, Egypt: contribution of schistosomiasis and hepatitis virus infections. *J. Hepatol.* **26**:236–243.
- Arthur, R. R., N. F. Hassan, M. Y. Abdallah, M. S. El Sharkawy, M. D. Saad, B. G. Hackbart, and I. Z. Imam. 1997. Hepatitis C antibody prevalence in blood donors in different governorates in Egypt. *Trans. R. Soc. Trop. Med. Hyg.* **91**: 271–274.
- Centers for Disease Control and Prevention. 1998. Morbidity and Mortality Weekly Report **47**:No. RR-19.
- Drummond, A. J., G. K. Nicholls, A. G. Rodrigo, and W. Solomon. 2002. Estimating mutation parameters, population history and genealogy simultaneously from temporally spaced sequence data. *Genetics* **161**:1307–1320.
- El-Zayadi, A., H. Abaza, S. Shawky, M. K. Mohamed, O. Selim, and H. M. Badran. 2001. Prevalence and epidemiological features of hepatocellular carcinoma in Egypt—a single centre experience. *Hepatol. Res.* **19**:170–179.
- El-Zayadi, A., O. Selim, S. Haddad, P. Simmonds, H. Hamdy, H. M. Badran, and S. Shawky. 1999. Combination treatment of interferon  $\alpha$ -2b and ribavirin in comparison to interferon monotherapy in treatment of chronic hepatitis C genotype 4 patients. *Ital. J. Gastroenterol. Hepatol.* **31**:472–475.
- El-Zayadi, A., P., Simmonds, H. Dabbous, L. Prescott, O. Selim, and A. Ahdy. 1996. Response to interferon- $\alpha$  of Egyptian patients infected with hepatitis C virus genotype 4. *J. Viral Hepatitis* **3**:261–264.
- Felsenstein, J. 1981. Evolutionary trees from DNA sequences: a maximum likelihood approach. *J. Mol. Evol.* **17**:368–376.
- Frank, C., M. K. Mohamed, G. T. Strickland et al. (12 co-authors). 2000. The role of parenteral antischistosomal therapy in the spread of hepatitis C in Egypt. *Lancet* **355**:887–891.
- Gad, A., E. Tanaka, K. Orii, A. Rokuhara, Z. Nooman, A. H. Serwah, M. Shoir, K. Yoshizawa, and K. Kiyosawa. 2001. Relationship between hepatitis C virus infection and schistosomal liver disease: not simply an additive effect. *J. Gastroenterol.* **36**:753–758.
- Grassly, N. C., P. H. Harvey, and E. C. Holmes. 1999. Population dynamics of HIV-1 inferred from gene sequences. *Genetics* **151**:427–438.
- Griffiths, R. C., and S. Tavaré. 1994. Sampling theory for neutral alleles in a varying environment. *Phil. Trans. R. Soc. Lond. Ser. B.* **344**:403–410.
- Habib, M., M. K. Mohamed, F. Abdel-Aziz, L. S. Magder et al. (12 co-authors) 2001. Hepatitis C virus infection in a community in the Nile Delta: risk factors for seropositivity. *Hepatology* **33**:248–253.
- Hassan, M. M., A. S. Zaghloul, H. B. El-Serag, O. Soliman, Y. Z. Patt, C. L. Chappell, R. P. Beasley, L. Y. Hwang. 2001. The role of hepatitis C in hepatocellular carcinoma—a case control study among Egyptian patients. *J. Clin. Gastroenterol.* **33**:123–126.
- Kalinina, O., H. Norder, S. Mukomolov, and L. O. Magnius. 2002. A natural intergenotypic recombinant of hepatitis C virus identified in St. Petersburg. *J. Virol.* **76**:4034–4043.
- Kamal, S. M., M. A. Madwar, T. Peters, R. Fawzy, and J. Rasenack. 2000. Interferon therapy in patients with chronic hepatitis C and schistosomiasis. *J. Hepatol.* **32**:172–174.
- Kao, J. H., and D. S. Chen. 2000. Transmission of hepatitis C virus in Asia: past and present perspectives. *J. Gastroenterol. Hepatol.* **15**:E91–E96.
- Kingman, J. F. C. 1982. On the genealogy of large populations. *J. Appl. Prob.* **19A**:27–43.
- Kuhner, M. K., J. Yamato, and J. Felsenstein. 1998. Maximum likelihood estimation of population growth rates based on the coalescent. *Genetics* **149**:429–434.
- Lande, R. 1998. Demographic stochasticity and Allee effect on a scale with isotropic noise. *Oikos* **83**:353–358.
- Maegraith, B. G. 1964. *Treatment of bilharziasis in Egypt*, UAR. World Health Organization, Geneva.
- May, R. M., S. Gupta, and A. R. McLean. 2001. Infectious disease dynamics: what characterises a successful invader? *Phil. Trans. R. Soc. Lond. Ser. B.* **356**:901–910.
- McCarthy, M. C., A. El Tigani, I. O. Khalid, and K. C. Hyams. 1994. Hepatitis B and C in Juba, southern Sudan: results of a serosurvey. *Trans. Roy. Soc. Trop. Med. Hyg.* **88**:534–536.
- Nafeh, M. A., A. Medhat, M. Shehata, N. N. Mikhail, Y. Swifee, M. Abdel-Hamid, S. Watts, A. D. Fix, G. T. Strickland, W. Anwar, and I. Sallam. 2000. Hepatitis C in a community in Upper Egypt: I. Cross-sectional survey. *Am. J. Trop. Med. Hyg.* **63**:236–241.
- Posada, D., and K. A. Crandall. 1998. Modeltest: testing the model of DNA substitution. *Bioinformatics* **14**:817–818.
- Pybus, O. G., M. A. Charleston, S. Gupta, A. Rambaut, E. C. Holmes, and P. H. Harvey. 2001. The epidemic behaviour of the hepatitis C virus. *Science* **292**:2323–2325.
- Pybus, O. G., A. Rambaut, and P. H. Harvey. 2000. An integrated framework for the inference of viral population history. *Genetics* **155**:1429–1437.
- Quinti, I., E. Renganathan, E. El Ghazzawi, M. Divizia, G. Sawaf, S. Awad, A. Pana, and G. Rocchi. 1995. Seroprevalence of HIV and HCV infections in Alexandria, Egypt. *Zentralbl. Bakteriol.* **283**:239–244.
- Rambaut, A. 2000. Estimating the rate of molecular evolution: incorporating non-contemporaneous sequences into maximum likelihood phylogenies. *Bioinformatics* **16**:395–399.
- Ray, S. C., R. R. Arthur, A. Carella, J. Bukh, and D. L. Thomas. 2000. Genetic epidemiology of hepatitis C virus throughout Egypt. *J. Infect. Dis.* **182**:698–707.
- Rodrigo, A. G., E. G. Shpaer, E. L. Delwart, A. K. Iversen, M. V. Gallo, J. Brojatsch, M. S. Hirsch, B. D. Walker, and J. I. Mullins. 1999. Coalescent estimates of HIV-1 generation time in vivo. *Proc. Natl. Acad. Sci. USA* **96**:2187–2191.
- Salemi, M., and A. M. Vandamme. 2002. Hepatitis C virus evolutionary patterns studied through analysis of full genome sequences. *J. Mol. Evol.* **54**:62–70.
- Schierup, M. H., and J. Hein. 2000. Consequences of recombination on traditional phylogenetic analysis. *Genetics* **156**:879–879.
- Shobokshi, O., E. Serebour, L. Skakni et al. 2002. Peg-IFN  $\alpha$ -2a as a monotherapy or in combination with ribavirin significantly improves end of treatment response rate in HCV genotype 4 chronic active hepatitis patients. Abstract. 2002 Meeting of the American Association for the Study of Liver Disease.
- Shobokshi, O., E. Serebour, L. Skakni, Y. Al Saffy, and M. Ahdal. 1999. Hepatitis C genotypes and subtypes in Saudi Arabia. *J. Med. Virol.* **58**:44–48.
- Smith, D. B., S. Pathirana, F. Davidson, E. Lawlor, P. L. Yap, and P. Simmonds. 1997. The origin of hepatitis C genotypes. *J. Gen. Virol.* **78**:321–328.

- Strimmer, K., and O. G. Pybus. 2001. Exploring the demographic history of DNA sequences using the generalized skyline plot. *Mol. Biol. Evol.* **18**:2298–2305.
- World Health Organization. 1997*a*. Hepatitis C. *Weekly Epidemiological Record* **72**:65–69.
- . 1997*b*. Hepatitis C: global prevalence. *Weekly Epidemiological Record* **72**:341–344.
- Worobey, M. 2001. A novel approach to detecting and measuring recombination: new insights into evolution in viruses, bacteria and mitochondria. *Mol. Biol. Evol.* **18**:1425–1434.

Keith Crandall, Associate Editor

Accepted November 5, 2002