Epidemic dynamics of two coexisting hepatitis C virus subtypes

Nuria Jiménez-Hernández,1 Manuela Torres-Puente,1 Maria Alma Bracho,1 Inmaculada García-Robles,1 Enrique Ortega,2 Juan del Olmo,3 Fernando Carnicer,4 Fernando González-Candelas1 and Andrés Moya1

1Instituto Cavanilles de Biodiversidad y Biología Evolutiva and Departamento de Genética, Universidad de Valencia, Spain
2Unidad de Enfermedades Infecciosas, Hospital General de Valencia, Spain
3Servicio de Medicina Interna, Hospital Clínico de Valencia, Spain
4Unidad de Hepatología, Hospital General de Alicante, Spain

Correspondence
Andrés Moya
andres.moya@uv.es

Received 6 June 2006
Accepted 18 September 2006

Hepatitis C virus (HCV) infection affects about 3 % of the human population. Phylogenetic analyses have grouped its variants into six major genotypes, which have a star-like distribution and several minor subtypes. The most abundant genotype in Europe is the so-called genotype 1, with two prevalent subtypes, 1a and 1b. In order to explain the higher prevalence of subtype 1b over 1a, a large-scale sequence analysis (100 virus clones) has been carried out over 25 patients of both subtypes in two regions of the HCV genome: one comprising hypervariable region 1 and another including the interferon sensitivity-determining region. Neither polymorphism analysis nor molecular variance analysis (attending to intra- and intersubtype differences, age, sex and previous history of antiviral treatment) was able to show any particular difference between subtypes that might account for their different prevalence. Only the demographic history of the populations carrying both subtypes and analysis of molecular variance (AMOVA) for risk practice suggested that the route of transmission may be the most important factor to explain the observed difference.

INTRODUCTION

Hepatitis C virus (HCV) is the only species of the genus Hepacivirus within the family Flaviviridae. It possesses a positive-sense, single-stranded RNA genome of about 9600 nt, characterized by a high level of heterogeneously distributed genetic variability. Six phylogenetically well-defined clusters, denoted genotypes, and about 50 subtypes have been defined (Simmonds et al., 1993; Simmonds, 2004). Although they share most basic biological features, there seem to be some differences in persistence and interaction with the host cell among the different HCV genotypes, with important consequences for future therapy and public-health measures (Simmonds, 2004). The prevalence of HCV infection has increased in recent decades as a consequence of its incorporation into efficient transmission networks, such as blood transfusion and, most notably, injection-drug usage (IDU) (Pybus et al., 2005; Thomson & Finch, 2005). As a consequence, the distribution of subtypes and genotypes has gone through, and is still undergoing, dramatic changes. The relative frequencies of HCV genotypes in different populations are not uniform, reflecting changes in population groups, age at time of infection and route of transmission. In industrialized nations, subtype 1a is generally found in younger individuals, sharing IDU being the main risk factor. In contrast, subtype 1b is common in older individuals with a history of blood transfusion (Pybus et al., 2005). After the development of more efficient methods for blood screening, new cases of HCV infection due to blood transfusion have fallen dramatically and the relative frequency of genotypes 1b and 2 compared with that of other genotypes is decreasing (Webster et al., 2000; Payan et al., 2005). This is possibly a reflection of changing modes of infection over the last 30 years and the background prevalence of different genotypes among different age cohorts.

In Spain, the prevalence of HCV is similar to that in other industrialized nations (Wasley & Alter, 2000; Domínguez et al., 2001), although the distribution of HCV genotypes presents some differences with respect to other European countries (López-Labrador et al., 1997). Several studies (Forns et al., 1996; López-Labrador et al., 1997) have reported a high incidence (up to 86 %) of subtype 1b in Spain, but more recent surveys (Ramos-Sánchez et al., 2003;
Moreno Planas et al. (2005) have documented an increase of subtype 1a (up to 23.7% in a cohort of 355 individuals), probably reflecting the high incidence (about 86%) of HCV seropositivity among Spanish injection-drug users (Bolumar et al., 1996; Santana Rodriguez et al., 1998).

The application of population genetics and evolutionary epidemiology to the analysis of virus genetic variability has proved to have many potential benefits and uses (Moya et al., 2004). Among these, the application of coalescence theory (Kingman, 1982; Hudson, 1991) to the epidemiology of HCV (Pybus et al., 2001, 2003, 2005) has enabled the reconstruction of the epidemic and demographic history of a population from a sample of contemporary gene sequences. These studies have already provided some insight into the evolution of different HCV genotypes on different scales, from worldwide (Pybus et al., 2001) to national (Tanaka et al., 2002, 2004; Pybus et al., 2003; Nakano et al., 2004), or in specific risk groups (Pybus et al., 2005). Interestingly, apart from differences among virus genotypes and virus genome regions being analysed, these studies have also revealed important differences among the different reference populations considered (Nakano et al., 2004).

Knowledge of the processes and patterns that drive the epidemiology and evolution of different virus types also permits more sound inferences on their short- and medium-term future evolution and, in the case of HCV infection, this may have important consequences for public-health policies (Tanaka et al., 2002). As HCV subtype 1b seems to be associated with advanced liver disease in Spain (Lopez-Labrador et al., 1997) and both subtypes 1a and 1b belong to the HCV group with the worst response to antiviral treatment, it is very interesting to characterize and compare the epidemic and evolutionary behaviour of these two genotypes in Spanish populations. This was the main goal of this study, for which we have used information from a large sequence dataset derived in the context of analysing the epidemic and evolutionary behaviour of these two HCV genotypes in Spanish populations. This was the main goal of this study, for which we have used information from a large sequence dataset derived in the context of analysing the epidemic and demographic history of a population from a sample of contemporary gene sequences. These studies have already provided some insight into the evolution of different HCV genotypes on different scales, from worldwide (Pybus et al., 2001) to national (Tanaka et al., 2002, 2004; Pybus et al., 2003; Nakano et al., 2004), or in specific risk groups (Pybus et al., 2005). Interestingly, apart from differences among virus genotypes and virus genome regions being analysed, these studies have also revealed important differences among the different reference populations considered (Nakano et al., 2004).

A first round of PCR was then carried out in a 100 μl volume containing 10 μl of the reverse transcription product, 0.2 μM each dNTP, 400 nM genomic primer, 400 nM antigenic primer and 1.25 U Pfu DNA polymerase (Promega). The primers used, detailed in Table 1 of Bracho et al. (2004) (unless specified below), were: for region E1–E2, subtype 1a: 1-Eg1, 1-Eg2 and 1-Ea; and for subtype 1b: 2-Eg1, 2-Eg2 and 2-Ea; for region NSSA, subtype 1a: 1-Ng1, 2-Ng1, 1-Ng2, 2-Ng2, 1-Na and 2-Na. Primers for region NSSA, subtype 1b were Ng1 (5’-TGGAYGRGTRCGGYTGCAAGGTA), Ng2 (5’-CAGGTACGCTCRRGYRTGCA) and Na (5’-CCTCRAGGGGGCGCAT). In all cases, PCR was performed in a thermal cycler (Perkin Elmer 2400) as described previously (Bracho et al., 2004).

Cloning and sequencing of virus populations. Amplified DNA products for each region were purified and cloned directly into EcoRV-digested pBluescript II SK(+) phagemid (Stratagene). Cloned products for the E1–E2 region or NSSA region were sequenced by using vector-based primers KS and SK (Stratagene). For the former region, we obtained about 100 clones from each patient, yielding a total of 5060 sequences. For the NSSA region, we obtained between 25 and 96 clones per patient, and 3176 sequences in total were determined (see Supplementary Tables S3–S6, available in JGV Online).

Sequencing was carried out by using the ABI PRISM BigDye Terminator v3.0 system (Applied Biosystems) on an ABI 3700 automated sequencer. Sequences were verified and both strands were assembled by using the STADEN package (Staden et al., 2000). HCV sequences obtained in this study have been deposited in GenBank with accession nos AM271041–AM275326 and AM279768–AM282548 for regions E1–E2 and NSSA, respectively.

Sequence and polymorphism analysis. Sequence alignments were obtained by using CLUSTAL X v1.81 (Thompson et al., 1997). Several parameter estimates of DNA polymorphism were obtained with DnaSP 4.0 (Rozas et al., 2003). The following measures of genetic variability were estimated for virus samples of each patient: total number of mutations, number of segregating sites, haplotype and nucleotide diversity and number of pairwise nucleotide differences. The mean value and SD for each subtype and region were calculated. Differences were tested by means of t-tests when
data followed a normal distribution or, as in the case of haplotype diversity, by means of a non-parametric Mann–Whitney U-test. Additionally, genetic-diversity (intrapatient) and -differentiation (between patients) estimates were obtained by using MEGA 3.1 (Kumar et al., 2004).

**AMOVA.** To investigate further the effect of different factors on the partitioning of genetic variation, hierarchical AMOVAs were performed by using ARLEQUIN v2.0 (Schneider et al., 2000). Total variance was partitioned among groups, among patients within groups and within patients. The variance components were used to compute fixation indices (Wright, 1951, 1965; Slatkin, 1991) and statistical significance was tested by means of a non-parametric permutation approach (based on 3000 replicates) as described by Excoffier et al. (1992). Groups were defined according to the following factors: intersubtype differences, age, sex, risk factor and IFN treatment before the samples were collected.

**Estimation of epidemic dynamics.** The demographic history of these virus genotypes in this population was estimated by using parametric and non-parametric methods, based on coalescent theory (Griffiths & Tavaré, 1994), for each of the four gene/genotype combinations studied. As this approach is heavily demanding on computational resources, it was necessary to select a subset of the total number of sequenced clones to obtain results in a reasonable time. We used the following three subsets: (i) one random sequence clone obtained from each patient, for a total of 25 sequences for each subtype and genome region; (ii) the same total number of sequences, but now using the ‘consensus’ sequence from each patient; and (iii) the two most genetically divergent sequences among those obtained from each patient, thus representing 50 sequences per subtype and genome region.

The estimation procedure involved several steps (Pybus et al., 2005). Once the sequences were aligned, MODELTEST 3.1 (Posada & Crandall, 1998) and PAUP* 4.0b10 (Swofford, 2001) were used to derive the evolutionary-substitution model that best explained the data according to the Akaike information criterion (Akaike, 1974). A maximum-likelihood (ML) tree was obtained by using the best evolutionary model and with the molecular clock enforced. To convert the estimated genealogy into a timescale of years, we used the estimated evolutionary rates in the two genome regions from a set of sequences obtained from serially sampled patients (M. Torres-Puente, N. Jiménez, M. A. Bracho, I. García-Robles, F. Carnicer, J. del Olmo, E. Ortega, A. Moya & F. González-Candelas, unpublished data), using the procedure implemented in TipDate (Rambaut, 2000).

Briefly, HCV isolates from nine non-responder patients (three of whom were also included in this report: A09 and G14, of subtype 1a, and C08 of subtype 1b) were determined to evolve according to a molecular clock in both genome regions when clones from two time points – before and after IFN plus ribavirin treatment, about 6 months apart – were compared. Finally, the epidemic history and the demographic parameters were estimated from the corresponding time-corrected genealogies. This estimation was performed with the program GENIE 3.0 (Pybus & Rambaut, 2002). For more details of this method, see Holmes et al. (1995), Pybus et al. (2000, 2005) and Pybus & Rambaut (2002).

**Basic reproductive number: R₀.** This parameter and its confidence interval were derived from GENIE. R₀ is defined as the mean number of secondary infections derived from an infected individual in a completely susceptible population. Only when R₀ > 1 can infection spread over the population. Following Pybus et al. (2001), \( R₀ = R₀ \times \frac{R₀ - 1}{R₀ - 1 + r} \), where \( r \), the rate of exponential population growth, is estimated from the population-growth model obtained from GENIE and \( D \) is an empirical value of the mean duration of infectiousness. For HCV, we have considered two values for \( D \): 15 and 25 years.

### RESULTS

**Polymorphism and differentiation analyses**

Genetic variability in the virus sequences obtained from each patient was estimated with different measures of polymorphism (see Supplementary Tables S3–S6, available in JGV Online). Mean values and SEM for the different parameter estimates by genome region and genotype are summarized in Table 1 and compared in Fig. 1. Genotype 1b was systematically more polymorphic than genotype 1a for all measures of polymorphism, with the exception of haplotype diversity, where the mean value in the E1–E2 region for subtype 1a was slightly higher than that for subtype 1b, although with no statistical significance. This result is due to a single patient (C13) who harbours two quite different population sequences in the NS5A region. These populations are very divergent and one of them is characterized by a 9 bp insertion, apart from differences in many other sites. As this individual belongs to the donor/transfusion category, this difference presented statistical significance when comparisons were performed for different risks factors (see below). In contrast to the E1–E2 region, the

---

Table 1. Summary of genetic-variability parameter estimates

<table>
<thead>
<tr>
<th></th>
<th>S</th>
<th>Eta</th>
<th>Nhap</th>
<th>HD</th>
<th>Pi</th>
<th>Theta(S)</th>
<th>k</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>E1–E2 1a</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>64.240</td>
<td>69.320</td>
<td>52.760</td>
<td>0.842</td>
<td>0.019</td>
<td>0.026</td>
<td>9.153</td>
</tr>
<tr>
<td>SD</td>
<td>37.437</td>
<td>41.681</td>
<td>25.618</td>
<td>0.231</td>
<td>0.014</td>
<td>0.015</td>
<td>6.814</td>
</tr>
<tr>
<td>Max</td>
<td>118</td>
<td>126</td>
<td>89</td>
<td>0.997</td>
<td>0.044</td>
<td>0.048</td>
<td>20.549</td>
</tr>
<tr>
<td>Min</td>
<td>3</td>
<td>3</td>
<td>4</td>
<td>0.059</td>
<td>0.000</td>
<td>0.001</td>
<td>0.060</td>
</tr>
<tr>
<td>Median</td>
<td>75</td>
<td>78</td>
<td>55</td>
<td>0.955</td>
<td>0.021</td>
<td>0.031</td>
<td>9.815</td>
</tr>
<tr>
<td><strong>E1–E2 1b</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>78.800</td>
<td>85.960</td>
<td>65.040</td>
<td>0.825</td>
<td>0.025</td>
<td>0.032</td>
<td>11.693</td>
</tr>
<tr>
<td>SD</td>
<td>44.092</td>
<td>49.375</td>
<td>32.745</td>
<td>0.321</td>
<td>0.019</td>
<td>0.018</td>
<td>8.935</td>
</tr>
<tr>
<td>Max</td>
<td>150</td>
<td>166</td>
<td>97</td>
<td>0.999</td>
<td>0.058</td>
<td>0.061</td>
<td>27.330</td>
</tr>
<tr>
<td>Min</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>0.039</td>
<td>0.000</td>
<td>0.001</td>
<td>0.039</td>
</tr>
<tr>
<td>Median</td>
<td>87</td>
<td>94</td>
<td>78</td>
<td>0.993</td>
<td>0.025</td>
<td>0.036</td>
<td>11.863</td>
</tr>
<tr>
<td><strong>NS5A 1a</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>46.120</td>
<td>47.680</td>
<td>28.440</td>
<td>0.725</td>
<td>0.008</td>
<td>0.013</td>
<td>5.928</td>
</tr>
<tr>
<td>SD</td>
<td>39.233</td>
<td>41.145</td>
<td>19.264</td>
<td>0.341</td>
<td>0.008</td>
<td>0.011</td>
<td>5.582</td>
</tr>
<tr>
<td>Max</td>
<td>125</td>
<td>131</td>
<td>64</td>
<td>1.000</td>
<td>0.028</td>
<td>0.035</td>
<td>20.808</td>
</tr>
<tr>
<td>Min</td>
<td>3</td>
<td>3</td>
<td>4</td>
<td>0.066</td>
<td>0.000</td>
<td>0.001</td>
<td>0.081</td>
</tr>
<tr>
<td>Median</td>
<td>30</td>
<td>31</td>
<td>28</td>
<td>0.928</td>
<td>0.005</td>
<td>0.008</td>
<td>3.489</td>
</tr>
<tr>
<td><strong>NS5A 1b</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>79.520</td>
<td>82.840</td>
<td>41.680</td>
<td>0.874</td>
<td>0.014</td>
<td>0.023</td>
<td>10.211</td>
</tr>
<tr>
<td>SD</td>
<td>52.590</td>
<td>55.534</td>
<td>19.437</td>
<td>0.248</td>
<td>0.011</td>
<td>0.015</td>
<td>8.441</td>
</tr>
<tr>
<td>Max</td>
<td>155</td>
<td>164</td>
<td>75</td>
<td>0.999</td>
<td>0.042</td>
<td>0.045</td>
<td>31.099</td>
</tr>
<tr>
<td>Min</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>0.070</td>
<td>0.000</td>
<td>0.001</td>
<td>0.070</td>
</tr>
<tr>
<td>Median</td>
<td>81</td>
<td>85</td>
<td>45</td>
<td>0.984</td>
<td>0.012</td>
<td>0.026</td>
<td>9.252</td>
</tr>
</tbody>
</table>
differences were always statistically significant ($P < 0.05$) for the NS5A region (Fig. 1).

A common feature for both regions and virus subtypes is the extraordinary heterogeneity observed for the different parameters estimated within every group. In each of these, some patients harboured highly variable virus populations, with parameters close to or even equal to the maximum possible value for the analysed sample, whereas others presented almost completely monomorphic populations. This pattern leads to large variances in the estimates of the different parameters and, in consequence, to very conservative tests for pairwise comparisons between groups. A possible solution to obtain a more realistic picture of the different underlying parameters is to compare the medians instead of the means (Table 1). This comparison again results in the same general conclusions already discussed.

Despite the high heterogeneity in genetic variability within patients, most genetic variation is found at the inter-patient level (Table 2; Supplementary Tables S7–S10 and Supplementary Fig. S1, available in JGV Online). This pattern is consistent for both genome regions and virus subtypes, with about 85% of the total genetic variation being found among patients. Nevertheless, there are several cases in which genetic differentiation between patients was lower than genetic diversity within patients. So, for subtype 1a, the pairwise differences for patients C14, C17 and C38 are lower than several intrapatient diversities for both the E1–E2 and NS5A regions. These three patients were known to be related epidemiologically, as they were infected from a common source in a recent, large HCV outbreak (F. González-Candelas, M. A. Bracho, B. Wróbel & A. Moya, unpublished data). In contrast, four HCV-1b-infected patients (A14, A35, C10 and C19) presented lower

![Fig. 1. Graphic representation of the mean values and SEM for several parameters of genetic variation, for both subtypes and regions. Asterisks indicate statistically significant differences between subtypes.](image)

<table>
<thead>
<tr>
<th>Subtype 1a</th>
<th>Subtype 1b</th>
</tr>
</thead>
<tbody>
<tr>
<td>E1–E2</td>
<td>NS5A</td>
</tr>
<tr>
<td>Mean within-patient diversity</td>
<td>0.0194</td>
</tr>
<tr>
<td>Mean among-patients diversity</td>
<td>0.0966</td>
</tr>
<tr>
<td>Mean overall diversity</td>
<td>0.1160</td>
</tr>
<tr>
<td>Proportion of interpopulation differentiation</td>
<td>0.8329</td>
</tr>
</tbody>
</table>
interpatient divergence in the NS5A region than several intrapatient diversities, but this was not the case for the E1–E2 region, although the intrapatient diversities for these patients in this region were among the lowest (see Supplementary Table S8, available in JGV Online). Three of these patients (A14, A35 and C10) had unknown risk factors for HCV infection (see Supplementary Table S2) and the fourth presumably became infected through IDU. Hence, a possible relationship among these patients (including also patient C15) cannot be ruled out. If it does exist, it is reflected only on a slowly evolving region of the genome, in which changes have not yet accumulated to such an extent as to delete every footprint of their common origin.

A more appropriate statistical methodology for testing the differential contribution of different factors to this partitioning of molecular variation is provided by AMOVA, which we performed next.

**AMOVA**

HCV genetic variation was analysed with regard to several factors that might generate significant differences among the groups considered: intersubtype differences, age and sex of the patients, previous treatment with IFN and risk factor for HCV infection.

When considering intersubtype differences, variance was partitioned among and within patients belonging to the same subtype for each genome region. Table 3 shows the corresponding AMOVAs. As can be observed for both regions, the highest percentage of variation is due to differences among subtypes, followed by differences among patients and within patients.

Patients were grouped by age into the five following categories: (i) between 20 and 30 years; (ii) between 31 and 40 years; (iii) between 41 and 50 years; (iv) between 51 and 60 years; and (v) between 61 and 70 years. We defined four groups for risk factor: (i) patients who acquired HCV by IDU; (ii) patients subjected to blood transfusion or blood donors; (iii) patients infected in an epidemic outbreak; and (iv) patients who acquired HCV from an unknown source (community-acquired HCV). For these factors, as well as for sex and previous administration of IFN, the total genetic variance was partitioned hierarchically among factors, among patients within factors and within patients. No significant differences were detected either for subtype or genome region for age (see Supplementary Tables S11 and S12, available in JGV Online), sex (see Supplementary Tables S13 and S14) or previous IFN treatment (see Supplementary Tables S15 and S16). However, significant differences appeared between groups defined by risk practice for genotype 1a in both regions (Table 4). These analyses were performed by considering three groups for subtype 1a: IDU, outbreak and of unknown origin, and for subtype 1b: IDU, donor/transfusion and of unknown origin. Similar results (not shown) were obtained when analyses were carried out excluding the ‘unknown-origin’ category, although with a tendency to increase the difference observed due to the effect of the risk factor. The lack of significance in the case of subtype 1b can be explained by the high genetic divergence within patients (see Fig. 1), probably resulting from its older presence in this population.

Except for the NS5A region in subtype 1b patients, virus sequences from IDU patients were genetically more diverse than those derived from an outbreak (subtype 1a) or from being donors or subjected to transfusion (subtype 1b, see Table 5). The exception is due to a single patient (C13, see Supplementary Table S6 in JGV Online) who harbours two quite different population sequences in the NS5A region. These populations are very divergent, resulting in a high value for the nucleotide-diversity estimate within this patient, and one of them is characterized by a 9 bp insertion, apart from differences in many other sites. Hence, as this is a unique event and, on a sample of only three patients in this category, it has a strong influence on the global mean for this group, we interpret this observation more as an effect of sampling than as a feature to be explained by the kind of transmission (donor/transfusion) common to patients from this group.

<table>
<thead>
<tr>
<th>Table 3. AMOVAs for differences among subtypes</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Region</strong></td>
</tr>
<tr>
<td>E1–E2</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>NS5A</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

*P < 0.05.
Table 4. AMOVAs for the risk groups corresponding to the E1–E2 and NS5A regions

<table>
<thead>
<tr>
<th>Region</th>
<th>Subtype</th>
<th>Source of variation</th>
<th>d.f.</th>
<th>Sum of squares</th>
<th>Variance components</th>
<th>Percentage of variation</th>
<th>Fixation indices</th>
</tr>
</thead>
<tbody>
<tr>
<td>E1–E2</td>
<td>1a</td>
<td>Among groups</td>
<td>2</td>
<td>8933.98</td>
<td>3.210 Va</td>
<td>10.92</td>
<td>0.109*</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Among patients within groups</td>
<td>20</td>
<td>44202.05</td>
<td>21.752 Vb</td>
<td>74.04</td>
<td>0.850*</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Within patients</td>
<td>2310</td>
<td>10207.05</td>
<td>4.419 Vc</td>
<td>15.04</td>
<td>0.831*</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Total</td>
<td>2332</td>
<td>63343.07</td>
<td>29.380</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1b</td>
<td>Among groups</td>
<td>2</td>
<td>6164.33</td>
<td>−0.237 Va</td>
<td>−0.63</td>
<td>−0.006**</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Among patients within groups</td>
<td>18</td>
<td>58036.96</td>
<td>32.199 Vb</td>
<td>85.02</td>
<td>0.844*</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Within patients</td>
<td>2076</td>
<td>12268.17</td>
<td>5.910 Vc</td>
<td>15.60</td>
<td>0.845*</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Total</td>
<td>2096</td>
<td>76469.47</td>
<td>37.871</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NS5A</td>
<td>1a</td>
<td>Among groups</td>
<td>2</td>
<td>5124.66</td>
<td>1.938 Va</td>
<td>6.52</td>
<td>0.065*</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Among patients within groups</td>
<td>20</td>
<td>31008.86</td>
<td>25.346 Vb</td>
<td>85.29</td>
<td>0.918*</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Within patients</td>
<td>1407</td>
<td>3424.80</td>
<td>2.434 Vc</td>
<td>8.19</td>
<td>0.912*</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Total</td>
<td>1429</td>
<td>39558.32</td>
<td>29.719</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1b</td>
<td>Among groups</td>
<td>2</td>
<td>5031.74</td>
<td>1.408 Va</td>
<td>3.74</td>
<td>0.037**</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Among patients within groups</td>
<td>18</td>
<td>37089.98</td>
<td>30.860 Vb</td>
<td>81.93</td>
<td>0.857*</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Within patients</td>
<td>1382</td>
<td>7460.49</td>
<td>5.398 Vc</td>
<td>14.33</td>
<td>0.851*</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Total</td>
<td>1402</td>
<td>49582.20</td>
<td>37.666</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*P<0.05; **ns, not significant.

Demographic history of subtypes 1a and 1b and the basic reproductive number

We have also investigated the demographic history of both subtypes in order to discern underlying reasons for the higher prevalence of genotype 1b over 1a in this population. Fig. 2 shows the ML estimates of effective population sizes through time and the corresponding non-parametric estimates (i.e. generalized skyline plots) for both genotypes and genome regions analysed. The graphical representation of the epidemic history is scaled in years. As evolutionary rates, we used the following estimates: 3.33 × 10^{-3} and 1.75 × 10^{-3} substitutions per site per year for the E1–E2 and NS5A regions, respectively (M. Torres-Puente, N. Jiménez, M. A. Bracho, I. García-Robles, F. Carnicer, J. del Olmo, E. Ortega, A. Moya & F. González-Candelas, unpublished data).

Table 5. Mean nucleotide diversity (±SD) for the different risk groups

<table>
<thead>
<tr>
<th>Subtype</th>
<th>Risk group</th>
<th>No. patients</th>
<th>E1–E2 region</th>
<th>NS5A region</th>
</tr>
</thead>
<tbody>
<tr>
<td>1b</td>
<td>IDU</td>
<td>3</td>
<td>0.02313 ± 0.01797</td>
<td>0.01245 ± 0.01064</td>
</tr>
<tr>
<td></td>
<td>Donor/transfusion</td>
<td>3</td>
<td>0.00647 ± 0.00539</td>
<td>0.01556 ± 0.02279</td>
</tr>
<tr>
<td></td>
<td>Unknown</td>
<td>15</td>
<td>0.02899 ± 0.01914</td>
<td>0.01496 ± 0.01018</td>
</tr>
<tr>
<td>1a</td>
<td>IDU</td>
<td>11</td>
<td>0.02264 ± 0.01651</td>
<td>0.00966 ± 0.00695</td>
</tr>
<tr>
<td></td>
<td>Outbreak</td>
<td>3</td>
<td>0.00605 ± 0.00179</td>
<td>0.00137 ± 0.00166</td>
</tr>
<tr>
<td></td>
<td>Unknown</td>
<td>9</td>
<td>0.01824 ± 0.01364</td>
<td>0.00534 ± 0.00517</td>
</tr>
</tbody>
</table>

The demographic models with the highest likelihood varied among datasets (Fig. 2). The datasets of genotype 1a fitted a model of exponential growth and piecewise demographic expansion in the NS5A and E1–E2 regions, respectively. The latter is a two-phase model of population growth, consisting of an initial period of constant population size followed by a period of exponential growth. The difference in the ML models for both regions is probably a reflection of the different best substitution-rate models. Nevertheless, both models indicated an increase in growth rate and onset of exponential growth during the 1980s. The panorama offered by genotype 1b was different. In this case, we observed that the epidemic-growth rate slowed down over time, reaching a steady state. This behaviour was reflected in a logistic-growth model for both regions (Fig. 2). The number of newly infected individuals with this genotype appears to
have increased slowly or stalled during the early 1980s. Noticeably, this analysis shows a time overlap in the change of epidemic-growth phase in the two virus genotypes, with an increase in the growth rate of subtype 1a corresponding to the decrease of 1b. It should be indicated that similar results were obtained when the input dataset was not a random sequence for each patient, but the two most divergent or the consensus sequence.

Finally, we have estimated $R_0$, the basic reproductive number, for both subtypes, assuming mean times of infectiousness of 15 and 25 years. Table 6 shows the estimates of $R_0$, with confidence intervals (except in one case where it was not possible), according to the three applied methods used to sample sequences from patients. As can be observed, firstly, $R_0$ is always $>1$, and second, when comparing subtypes 1a and 1b, both regions always show the same pattern. However, there is an effect on $R_0$ of the sampling method of sequences from patients, because such a pattern shows the opposite direction when sampling was done on the two most divergent sequences compared with the other two. It is worth noticing, however, that the random sequence or the consensus sequence yields closer estimates. Larger values for $R_0$ were obtained when using $D = 25$ instead of $D = 15$, but the general pattern just described was the same regardless of the time considered for infectiousness.

**DISCUSSION**

HCV genotypes and subtypes exhibit a complex epidemiological pattern with respect to geographical distribution, prevalence, response to treatment and transmission mode. Ascertaining the factors responsible for such complex epidemiology is a difficult task and, if solved, it can contribute to act on the disease. Although studies (Mondelli

---

**Fig. 2.** Coalescence-based estimates of demographic dynamics of HCV subtypes 1a and 1b in the studied population. ML parametric (continuous lines) and non-parametric (skyline plots; dotted lines) estimates are shown for each genome region (upper panel, E1–E2; lower panel, NS5A). The $x$ axes represent time in years; the $y$ axes indicate effective population size.

**Table 6.** Estimates of $R_0$ (with 95% confidence intervals) for both subtypes and regions according to the three applied methods of sampling sequences from patients

Two times of infectiousness ($D = 15$ and $D = 25$ years) were considered.

<table>
<thead>
<tr>
<th>Method</th>
<th>Region</th>
<th>Subtype</th>
<th>$R_0$ ($D = 15$)</th>
<th>$R_0$ ($D = 25$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Two most divergent</td>
<td>E1–E2</td>
<td>1a</td>
<td>2.782 (2.406–3.131)</td>
<td>2.970 (2.343–3.552)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1b</td>
<td>4.093 (2.998–5.550)</td>
<td>5.155 (3.330–7.583)</td>
</tr>
<tr>
<td></td>
<td>NSS5A</td>
<td>1a</td>
<td>2.565 (2.129–2.963)</td>
<td>2.608 (1.882–3.272)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1b</td>
<td>5.774 (4.418–7.460)</td>
<td>7.957 (5.697–10.767)</td>
</tr>
<tr>
<td>Consensus</td>
<td>E1–E2</td>
<td>1a</td>
<td>5.139 (3.792–6.549)</td>
<td>6.898 (4.653–9.248)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1b</td>
<td>3.276 (2.739–3.841)</td>
<td>3.793 (2.898–4.735)</td>
</tr>
<tr>
<td></td>
<td>NSS5A</td>
<td>1a</td>
<td>4.474</td>
<td>5.790</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1b</td>
<td>3.400 (2.805–3.987)</td>
<td>4.000 (3.008–4.978)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1b</td>
<td>2.970 (2.498–3.471)</td>
<td>3.283 (2.497–4.118)</td>
</tr>
<tr>
<td></td>
<td>NSS5A</td>
<td>1a</td>
<td>4.468 (3.132–6.507)</td>
<td>5.780 (3.553–9.178)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1b</td>
<td>3.744 (3.045–4.489)</td>
<td>4.573 (3.408–5.815)</td>
</tr>
</tbody>
</table>
& Silini, 1999; Zein, 2000) have reported that genetic
differences between virus subtypes 1a and 1b could account
for their prevalence and clinical outcome, there are no
conclusive results yet. In this work, we have pursued a
similar objective, which we approached by studying the
effect of several factors on partitioning the observed genetic
variation in both virus subtypes. In addition, we have made
use of an analytical tool based on coalescent theory (Pybus
et al., 2000; Strimmer & Pybus, 2001) to make inferences on
the epidemic history of both HCV subtypes.

Subtype 1b is more polymorphic than 1a (Fig. 1) in our
population, and also when it is grouped into smaller groups
according to risk factor (Table 4). This observation has been
also reported elsewhere (Tanaka et al., 2002; Simmonds,
2004). Our results indicate that this might be due to subtype
1b being established in this population for far longer than
subtype 1a.

The E1–E2 region is more polymorphic than NS5A and the
different measures of genetic polymorphism used showed
statistically significant differences between both subtypes for
region NS5A, but not for E1–E2. It seems that selective
constraints are acting differently on both regions
(unpublished data). The samples used in this analysis are a
subset of those used in a larger study on the effects of IFN
and ribavirin treatment on the patients’ response and virus
genetic variation. The two genome regions analysed in this
study were chosen because they seem to be associated with
the virus response to treatment. So, the NS5A gene contains
the ISDR and the E1–E2 gene fragment analysed encom-
passes HVR1 and HVR2, which might account for their
different evolutionary behaviour. Although it is usually the
NS5B gene that has been used for HCV genotyping and
molecular-epidemiology studies, the epidemic dynamics
analysed here should be reflected similarly throughout the
whole HCV genome, provided that the region under analysis
accumulates enough differences in the time interval
considered. This is the case for both regions, although
their different evolutionary rates and possibly different
selective regimes translate into the reported differences.

The results obtained with respect to the demographic
history support the hypothesis that the differences in
prevalence between subtypes could be due to differences in
the route of transmission, which also correlate with the
frequency of such routes through time (Mortimer, 1995;
Smith et al., 1997; Webster et al., 2000; Zein, 2000). These
studies have established a relationship between route of
transmission and virus genotypes and subtypes. Specifically,
it has been shown that genotype 1a is associated
preferentially with young intravenous-drug users; mean-
while, genotype 1b is more abundant in people over 50 years
of age and commonly associated with blood transfusion and
sporadic hepatitis (Pawlotsky et al., 1995; Pol et al., 1995;
Simmonds et al., 1996). Accordingly, and with an argument
similar to that used previously for other populations
(Kalina et al., 2001; Yu et al., 2001), we could hypothesize
a change in prevalence of subtype 1a over 1b in the near
future in this Spanish population, with corresponding
changes in the incidence of disease complications associated
differentially with the subtypes (Tanaka et al., 2002).

Hepatitis C is a young disease for science, as its causative
agent was unknown until 1989 (Choo et al., 1989). For this
reason, until then, no methods were available to screen
blood successfully for HCV and to implement other
measures to prevent its infectious spread. Around the
early 1980s, the existence of a nonA–nonB hepatitis virus
was known and there were also a few methods to prevent
virus transmission by contaminated blood products, which
diminished the incidence of subtype 1b slightly until the
discovery of the virus in 1989 (Memon & Memon, 2002).
After its discovery, new cases of infection due to blood
transfusion have fallen dramatically (Donahue et al., 1992),
which would explain both the growth-rate stabilization of
subtype 1b and its higher prevalence in patients over
50 years of age.

The demographic dynamics shown in Fig. 2 indicate that
subtype 1b epidemics appeared before those of 1a. The
population dynamics inferred from coalescent theory for
subtype 1b are typical of a logistic growth, with an initial
phase of unknown duration of slow growth, followed by a
well-defined phase of exponential growth and, finally, a
steady state. Analysis of subtype 1a reveals a similar pattern,
but with a time lag of several decades. The exponential-
growth phase of subtype 1a started a few years after subtype
1b reached its stationary phase. This is an independent
indication that the dynamics of both subtypes have been
largely independent from each other, as they share neither
the main route of transmission nor the infected population.
Consequently, we would expect that subtype 1b was more
prevalent than 1a a few years ago, a situation likely to change
soon. Several reports from Spanish populations provide
empirical support for this prediction from our sequence-
based analysis (Forns et al., 1996; López-Labrador et al.,
1997; Ramos-Sánchez et al., 2003; Moreno Planas et al.,
2005).

This demographic pattern also predicts a change in
prevalence according to age. This is precisely what we
have observed. In the range of 20–30 years of age, 38.5 %
of patients were infected with subtype 1b, meanwhile 61.5 %
were infected with 1a; in the range of 30–40 years of age,
41.2 % were infected with 1b and 58.8 % with 1a; in the
range of 41–50 years of age, 69.2 % were infected with 1b
and 30.8 % with 1a; in the range of 51–60 years of age,
86.6 % were infected with 1b and 13.4 % with 1a; and finally,
in the range of 60–70 years of age, 100 % were infected with
1b. Why has subtype 1b changed its growth dynamics? It is
known that IDU has risen during recent years and a
proportionately high number of newly infected patients
acquired HCV through this route (Pawlotsky et al., 1995;
Webster et al., 2000). Due to the methods for HCV screening
in blood, infection through transfusions or blood donations
is very difficult and, for this reason, it is expected that, from
1989 onwards, most new cases of infection of HCV are

N. Jiménez-Hernández and others

Journal of General Virology 88
associated preferentially with IDU and other uncontrolled routes involving contaminated instruments or blood. This pattern accounts well for the recent increase in the growth rate found in the demographic history of subtype 1a, the higher prevalence among younger people and the decrease in growth rate for 1b. The temporal coincidence in the rise of subtype 1a seems to be related more directly to the increase of IDU in Western countries rather than to a competitive displacement of the previous prevalent subtype, HCV-1b, by a new variant with higher fitness.

Further support for this hypothesis derives from the AMOVA results on risk factor and the estimates of $R_0$ presented here. Risk factors affect subtypes 1a and 1b differentially (Table 4), where there is a higher divergence between IDU patients of subtype 1a with respect to the rest of the groups than between those of subtype 1b against their corresponding groups, which is in favour of a preferential transmission of 1a through IDU. Our estimates of $R_0$, the basic reproductive number, are in all cases $> 1$ (Table 6). This indicates that the epidemics will take off in both subtypes. However, there are disparities in the estimates of $R_0$ according to the sampling of sequences from patients, something that merits comment. When considering the consensus sequence or a sequence taken randomly, first, both estimates are similar for each subtype and genome region and, second, they point to a future imposition of subtype 1a over 1b. The result was completely the opposite when sampling was done by using the two most divergent sequences from each patient. In addition, the estimates are the most divergent among the three methods. This observation, together with evidence presented in this work favouring the exponential expansion of subtype 1a and stability of 1b, casts some doubts about the validity of sampling the two most divergent sequences. A higher $R_0$ value for subtype 1a is expected if we consider that effective measures to control IDU are lower than those applied to control HCV among donors or in transfusions, which also supports the replacement of subtype 1b by 1a.

Finally, we have used our own estimates for the evolutionary rates in the two regions analysed. The main reason is that these estimates were obtained from a sample of patients that partially overlapped those used in this analysis and for exactly the same genome regions considered here. None of the previously available estimates (Abe et al., 1992; Smith et al., 1997; Pybus et al., 2001; Cochrane & Simmonds, 2002) matched the limits used in this case exactly. This may contribute to explaining the difference in the values obtained in our estimates from those of other authors, whose estimates range between $0.7 \times 10^{-3}$ and $1.92 \times 10^{-3}$ for the E1–E2 genes and between $0.4 \times 10^{-3}$ and $0.5 \times 10^{-3}$ for the NS5 gene. There are, however, some other points to be considered. Firstly, even though we chose for the estimation procedure only a reduced subset of patients in whom the molecular clock could not be rejected in either of the two regions considered, we obtained highly variable estimates for the different patients, especially in the E1–E2 region, with differences of up to three orders of magnitude. This extraordinary variability is also reflected in the range of values for different parameter estimates of genetic polymorphism (Table 1; Supplementary Tables S3 and S4, available in JGV Online), although in this case, the time since each patient was infected before sampling certainly contributes to the differences observed. In consequence, it is very adventurous to assign a single rate of evolution to an organism such as HCV, which can exhibit such a wide variation depending on the niche (i.e. patient) that it colonizes. Secondly, our estimates were based on the analysis of changes accumulated over a relatively short time period, 6 months, as opposed to most other estimates in which the time involved has been usually measured in years. There is a well-known inverse relationship between evolutionary rate and time of evolution (Holmes, 2003) and it is not clear which estimates should be used in specific contexts, such as the reconstruction of epidemic behaviour, because although the period covered may expand for decades or even hundreds of years in the case of HCV, the actual sequences used in the analysis display a range of divergences, from months to years. In any case, our translation of the epidemic history into absolute time by using our own evolutionary rates matches very closely what is known about the recent epidemiology of these two HCV subtypes in the reference population.

ACKNOWLEDGEMENTS

This work was supported by Conselleria de Sanitat i Consum, Generalitat Valenciana (Spain) and project BFU2005-00503 from Ministerio de Educación y Ciencia (Spain).

REFERENCES


Prevalence of hepatitis C virus genotypes

Arrieta, F., Jimenez Garrido, M., Sanchez Turrion, V. & Cuervas-Boullosa Grana, E., Rubio Gonzalez, E., Martin Garcia, S., Martinez

Genitourin Med

J Hepatol

geno type 1b, cirrhosis and hepatocellular carcinoma. Spanish patients with HCV infection: relationship between HCV

J Clin Microbiol

from Spanish patients.

three methods for genotyping hepatitis C virus strains in samples


Hepatitis C: an

arsphenamine jaundice and the recognition


Payan, C., Roudot-Thoraval, F., Marcellin, P., Bled, N., Duvergie, G., Fouchard-Hubert, I., Trimoulet, P., Couzigou, P., Cointe, D. & other


