Infection, Genetics and Evolution 10 (2010) 1252–1261

Contents lists available at ScienceDirect

Infection, Genetics and Evolution

journal homepage: www.elsevier.com/locate/meegid

Relating the liver damage with hepatitis C virus polymorphism in core region and human variables in HIV-1-coinfected patients

Marina Matas [a], Antònia Picornell [a], Carmen Cifuentes [b], Antoni Payeras [b], Antoni Bassa [b], Francesc Homar [b], F. Xavier López-Labrador [c], Andrés Moya [c], Maria M. Ramon [a], José A. Castro [a,*]

[a] Institut Universitari d'Investigacions en Ciències de la Salut (IUNICS) i Laboratori de Genètica, Departament de Biologia, Universitat de les Illes Balears, Campus de la UIB, Palma de Mallorca, Illes Balears, 07122, Spain
[b] Unitat de Malalties Infeccioses, Hospital Son Llàtzer, Palma de Mallorca, Illes Balears, Spain
[c] Centre Superior d' Investigació en Salut Pública (CSISP), Av. de Catalunya 21, 46020 València; Institut Cavanilles de Biodiversitat i Biologia Evolutiva, Universitat de València, Apartado de Correos 22085, 46071 València, CIBER en Epidemiología y Salud Pública (CIBEResp), Instituto de Salud Carlos III, Spain

ARTICLE INFO

Article history:
Received 4 June 2010
Received in revised form 14 August 2010
Accepted 16 August 2010
Available online 21 August 2010

Keywords:
Hepatitis C virus
HCV
HIV
Coinfection
Generalized linear models
GLM

ABSTRACT

Hepatitis C virus (HCV) infection is the most important cause of chronic hepatitis, cirrhosis and end-stage liver disease leading to liver transplantation worldwide. Chronic infection by HCV causes liver fibrosis, which is accelerated by unknown mechanisms in patients with human immunodeficiency virus-1 (HIV-1) coinfection. Although the genetic variability of both HCV and HIV has been extensively studied in the context of monoinfections, more limited data is available regarding HCV–HIV coinfection. HCV disease progression among HIV coinfected patients may be influenced not only by demographic, epidemiological and clinical background variables, but also by genetic differences in infecting viruses. To explore this issue, we carried out a study in coinfected patients trying to associate the degree of liver damage to several demographic, clinical, and epidemiological characteristics of the patients, and also to the genetic variability of HCV between patients. For this purpose, we have applied different statistical techniques including the statistical generalized linear model (GLM) framework. The stage of fibrosis was indirectly measured by noninvasive means using the indexes Forns, APRI and FIB-4. HCV genetic variability between patients was estimated by sequencing the core region and by reconstructions of consensus maximum parsimony phylogenetic trees with 50% and 75% bootstrap majority rules. The results showed a direct correlation of the fibrosis biomarkers with the AST/ALT ratio, MoftIDU and with 3a HCV genotype clades, among others.

1. Introduction

Hepatitis C virus (HCV) is the only species of the genus *Hepacivirus* within the family *Flaviviridae*. HCV is a single-stranded, positive-sense RNA virus, with a genome of about 9.6 kb in length encoding for a polyprotein of about 3000 amino acids. HCV isolates are grouped in six phylogenetically well-defined clusters, denoted as genotypes, and about 50 subtypes (Simmonds et al., 2005). Although HCV is not considered to be cytophatic, the hallmark pathogenic effect of chronic infection is the generation of liver fibrosis with time (McCaughan and George, 2004). Fibrosis progression ultimately leads to cirrhosis of the liver and consequently to end-stage liver disease. The response to current standard-of-care HCV therapy is influenced by the HCV genotype: patients infected with HCV genotypes 1 or 4 show significantly lower sustained response rates than those infected with genotypes 2 or 3 (Heathcote, 2007). The viral polyprotein is processed by host and viral proteases to release three structural and seven non-structural proteins (Lindenbach and Rice, 2005). Two of the most studied HCV genome domains are the core and the NS5B coding regions. The core protein is an RNA-binding protein which forms the viral nucleocapsid, which interacts with cellular proteins, influencing numerous host cell functions (Lindenbach and Rice, 2005).

The prevalence of HCV positivity range between <1% in Northern Europe to more than 2% in Northern Africa, with the highest prevalence reported in Egypt (15–20%) (Alter, 2007). Since the development of more efficient methods for blood screening, new cases of HCV infection due to blood transfusion are extremely rare. The prevalence of HCV infection has increased in recent decades as a consequence of virus spread through efficient transmission networks, most notably injection drug use (IDU) (Pybus et al., 2005;

* Corresponding author at: Laboratori de Genètica, Departament de Biologia, Facultat de Ciències, Edifici Guillem Colom, Universitat de les Illes Balears, Crta Valldemossa Km. 7.5, 07122, Palma, Illes Balears, Spain. Tel.: +34 971 173153; fax: +34 971 173184.
E-mail address: jose.castro@uib.es (J.A. Castro).

1567-1348/$ – see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.meegid.2010.08.009

Thomson and Finch, 2005). 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. Consequently, the distribution of genotypes and subtypes has been changing over time. In industrialized nations, subtype 1a is generally found in younger individuals, sharing IDU as main risk factor. Otherwise, subtype 1b is more common in older individuals with a history of blood transfusion (Pybus et al., 2005). In Spain, the prevalence of HCV is similar to other industrialized nations, with HCV genotype 1 as the most prevalent followed by the genotype 3 (Esteban et al., 2008; Bruguera and Forns, 2006), although the distribution of HCV subtypes presents some differences from other European countries (Lopez-Labrador et al., 1997), with a higher incidence of subtype 1b infections. The distribution in our province of Mallorca is similar to that of the rest of the country (Cifuentes et al., 2004).

More recent study cohorts in Spain documented a decreasing relative frequency of subtype 1b and genotype 2 compared with an increase of subtype 1a infections, probably reflecting increasing incidence of subtype 1a infections among Spanish IDU (Jimenez-Hernandez et al., 2007). Due to common transmission routes, while the overall prevalence of HCV coinfection in individuals infected with human immunodeficiency virus-1 (HIV, *Retroviridae* family, genus *Lentivirus*) is around 2%, in HIV-positive patients with history of IDU the prevalence of HCV–HIV coinfection can reach up to 70–95% (Alter, 2006). HCV–HIV coinfection is a clinical problem worldwide, and is associated with a higher mortality than monoinfection with either virus alone (Rotman and Liang, 2009). Genotype 4, prevalent in the Middle East and different parts of Africa, is increasing among southern European intravenous drug users infected with HIV-1 (Franco et al., 2007). Two separate epidemics of HCV seem to have ocurred in Spain during the last 30 years. Firstly, one involved the spread of HCV genotypes 1a and 3. Latter, a more recent epidemic involved the spread of genotype 4 (Echevarría et al., 2006). Available data about the impact of HIV on HCV infection strongly suggests that coinfected patients have more risk of liver disease and a more rapid disease progression, as compared to HCV monoinfection (Matthews and Dore, 2008). In a prospective study in a Spanish cohort of HIV/HCV-coinfected patients, estimations over a period of time of 3 years determined that up to 44% of coinfected patients progressed one or more stages of fibrosis in such short period of time (Macias et al., 2009). Indeed, progression of HCV disease is accelerated in HIV–HCV coinfection, being more pronounced in patients with lower CD4+ cell counts (Reiberger et al., 2010). Whether variations in disease outcomes in immunosuppressed individuals are related to genetic differences in HCV strains is not well established (Gigou et al., 2001). In cohort study of HCV subtype 1b-infected patients, there is a correlation between HCV genetic similarity in core and NS5b genes and similarity of fibrosis progression in non-immunosuppressed patients, but not in liver-transplanted immunosupressed patients (Lopez-Labrador et al., 2004, 2006). Concurrently, there is evidence of decreased genetic diversity of HCV in HIV–HCV coinfection, compared to that in monoinfection, suggestive of reduced immune selective pressure (Lopez-Labrador et al., 2007; Shuhart et al., 2006).

The identification of the demographic and clinical variables predicting the clinical outcome of HCV–HIV coinfection represents one important challenge. Candidate factors include both viral (HIV and HCV viremia, HCV genotype, highly active antiretroviral therapy (HAART)) and host variables (age, sex, AST, ALT, albumin, platelet count, etc.), including some potentially involved in liver fibrosis (Nagayama et al., 2000; Walters et al., 2006; Bataller et al., 2003). The development of noninvasive tests indexes (Forns, APRI and FIB-4) as surrogate markers of liver fibrosis allow good estimations of liver damage; i.e. by using Class II fibrosis biomarkers which comprise a wide variety of biochemical scores and multi-parameter combinations. These indexes are based in a panel of standard laboratory tests of markers subject to variation in the serum or plasma of fibrotic patients, though only partially related to the mechanism of fibrogenesis (Gressner et al., 2007).

In the current study, we analysed the correlation between several demographic, clinical, and epidemiological characteristics of HCV–HIV-infected patients, and the HCV genetic sequence in the core region, with the degree of liver damage estimated by using noninvasive fibrosis measures. To evaluate the impact of the above-mentioned traits we applied an statistical approach based on generalized linear models (GLMs) (Venables and Ripley, 1999), to identify factors associated with liver disease progression.

2. Materials and methods

2.1. Patients

Thirty-seven HCV–HIV coinfected patients were chosen retrospectively from those followed in the HIV clinic of Son Llàtzer Hospital, Palma, Illes Balears (Spain), and the institutional review board approved the study. Inclusion criteria were prior informed consent to use biological samples for medical research, documented seropositivity and positive viremia for both HCV and HIV. No patient was coinfected with the hepatitis B virus (HBV). Stored serum or plasma samples ($-70\ ^{\circ}C$) were used for the study.

2.2. Variables

The following demographic, epidemiological, medical, laboratory, and substance use histories were obtained by chart review and recorded in detail when possible: sex, age, risk factors for HIV and HCV infection (including date of initial exposure if possible), HIV disease staging (1993 revised CDC classification), presence and compounds of HAART, alcohol consumption (alcohol abuse was defined by an average daily consumption >50 g for >2 years), HCV genotype, routine laboratory tests including CD4-positive (CD4+), cell counts and HIV and HCV viral loads. HIV disease staging was divided into two categories C3 and non-C3 (C3 and NC3 in Appendix A), due to the few occurrence of each stage, apart from C3 stage. HIV viral load was considered as low (L) when <1000 copies/mL and high (H) when >1000 copies/mL. Biochemical markers included: serum triglycerides (TGC), serum alanine-aminotransferase (ALT), and serum aspartate-aminotransferase (AST), glutamil-gamma transpeptidase (GGT) among others. HCV genotyping was performed using the VERSANT HCV Genotype 2.0 (LIPA) (Siemens); and HIV and HCV RNA quantification was performed using the COBAS AmpliPrep/COBAS TaqMan HIV-1 Test, versión 2.0. (Roche), and COBAS AmpliPrep/COBAS TaqMan HCV (Roche), CD4+ T cell counts were determined by standard flow cytometry and given in both absolute and relative numbers.

2.3. Outcome assessment: liver fibrosis

The degree of liver damage was measured by means of the indexes IF, APRI and FIB-4 (Forns et al., 2002; Sterling et al., 2006; Vallet-Pichard et al., 2007), due to the retrospective nature of the study and the lack of liver biopsies. The three indexes were used as response variables in the statistical modelling. We also analysed the three response variables together (simultaneously), to account for the existing correlation between indices obtained from the same individual.

The **IF** (Forns index) was calculated as described (Forns et al., 2002), by using the formula:

$$\mathrm{IF} = [7.811 - 3.131 \times \ln(\text{platelet count}) + 0.781 \times \ln(\mathrm{GGT}) + 3.467 \times \ln(\text{age}) - 0.014 \times (\text{cholesterol})]$$

The *APRI* (AST to platelet ratio) index was calculated as described (Wai et al., 2003) by using the formula:

$$\mathrm{APRI} = \left[\frac{(\text{AST level}(\mathrm{U/L}))/\mathrm{ULN}(\mathrm{U/L})}{\text{Platelet count}\,(10^9/\mathrm{L})}\right] \times 100$$

Being ULN the upper limit of normal AST level, we have considered this as 40.

The **FIB-4** index was calculated as described (Sterling et al., 2006), using the formula:

$$\text{FIB-4} = \text{age(yr)} \times \left[\frac{\mathrm{AST(U/L)}}{\text{Platelet count}(10^9/\mathrm{L}) \times \mathrm{ALT}(\mathrm{U/L})^{1/2}}\right]$$

2.4. HCV sequencing and phylogeny of the core region

Viral RNA was extracted from 400 μL of serum or plasma conserved at −70 °C with the High pure Viral RNA kit (Roche diagnostics, Barcelona, Spain), reverse transcription and PCR amplification of the 395 nucleotides from the core region (nucleotide 342–736 in the reference sequence AF009606) were performed following the method described by Ohno et al. (1997) with little modifications. Amplified products were purified by means of column affinity with the MSB Spin PCRapace kit (Invitek, Berlin, Germany) and direct bidirectional sequencing was done with the BigDye® Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems, Foster City, CA) and analysed in a 3130 Genetic Analyser (Applied Biosystems). All chromatograms were verified visually, and both strands assembled using Bioedit v.7.0.9.0 (Hall, 1999). HCV subtype assignation was confirmed by blast searches using the interface of the *Los Alamos HCV sequence and immunology database* (Kuiken et al., 2008).

Phylogenetic trees were constructed using MEGA (Tamura et al., 2007), by means of the maximum parsimony criterion with bootstrap support with 2000 replicates. We performed a close-neighbour-interchange (CNI) with search level 1. The initial tree for CNI search was random addition tree, with 10 replications. We used the consensus tree, based on either 50% or 75% majority rules, which were edited thereafter with the same MEGA software. To analyse the nucleotide diversity in the core region, we used the DNAsp software (Librado and Rozas, 2009).

2.5. Statistical methods

We applied a GLM framework (Venables and Ripley, 1999) using the free software *The R project for Statistical Computing* (http://www.r-project.org). Generalized linear modelling is a development of linear models to accommodate both non-normal response distributions and transformations to linearity in a clean and straightforward way. A generalized linear model may be described in terms of the following sequence assumptions: (i) there is a response, y, of interest and predictor variables x_1, x_2, . . ., whose values influence the distribution of the response and (ii) the predictor variables influence the distribution of y through a single *linear function, only*. This linear function is called the *linear predictor*, and is usually written:

$$y = \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_p x_p$$

The class of generalized linear models handled by the *R* package includes Gaussian, Binomial, Poisson, Inverse Gaussian and Gamma response distributions (families) and also Quasi-likelihood models where the response distribution is not explicitly specified. Each response distribution admits a variety of link functions to connect the mean with the linear predictor. Combining the "identity" link with the Gaussian family produces the normal linear model, with link function: $\beta x = \mu$ and mean function: $\mu = \beta x$. The maximum-likelihood estimates for this model are the ordinary least-squares estimates. When several response variables are calculated and used simultaneously in the analysis, with the same predictor variables, the combination of Gaussian family with "identity" link gives a multivariate normal analysis.

The Akaike's information criterion (*AIC*) was used to evaluate the best of the different models (lowest *AIC*). The *AIC* is known in the statistics trade as a *penalized log-likelihood*. In a model for which a log-likelihood value can be obtained, then

$$AIC = -2\,x\,\text{log-likelihood} + 2(p+1)$$

where p is the number of parameters in the model, and 1 is added for the estimated variance. *AIC* is useful because it explicitly penalizes any superfluous parameters in the model, by adding $2(p+1)$ to the deviance. When comparing two models, the smaller the *AIC*, the better the fit. This is the basis of automated model simplification using the function *Step* in the *R* (Crawley, 2007).

Finally, comparisons between different markers and other variables were done using SPSS for Windows, Rel. 15.0.1. 2006. Chicago: SPSS Inc. Data are expressed as percentages for categorical variables, and means and standard deviations for continuous variables.

3. Results

3.1. Baseline characteristics of study subjects

Table 1 shows the main features of the studied cohort, including the demographic and virological characteristics of the HIV–HCV-coinfected subjects and the variables considered as fibrosis predictors for the total cohort ($n = 37$). As can be seen, the distribution of the patients in the cohort is not homogenous. Most patients are males, with injection drug usage as method of transmission, without alcohol intake, with HAART and in a C3 HIV stage. The mean age of the patients is around 40 years with a mean duration of infection of 16 years. Table 2 indicates the same variables separated by the HCV infecting genotype.

3.2. HCV core phylogeny

Fig. 1 represents the maximum parsimony consensus phylogenetic tree for the core region. Two types of clusters could be differentiated depending on the majority rule cut-off value: a 50% or a more restrictive value of 75%. Using the 50% criteria, viral isolates were classified in 12 clades (A–L), whereas using the 75% criteria only five different clades were evident (a–e), which corresponded to the viral subtypes. Each of these clades (A–L and a–e) was introduced in the dataset to be included in the statistical modelling (Appendix A).

Table 1
Baseline demographic and viral characteristics of HIV–HCV-coinfected subjects and variables considered as fibrosis predictors. Number of patients, mean and standard deviation (SD) are indicated in the variables.

Characteristic	$n = 37$
Sex	
Male (M)	29
Female (F)	8
Method of transmission (Moft)	
Intravenous drug use (IDU)	34
Heterosexual contact (HET)	3
Alcohol consumption (ENO)	
Yes (Y)	8
No (N)	29
Highly active antiretroviral therapy (HAART)	
Yes	34
No	3
AGE, mean (SD), years	39.81 (6.01)
CD4 cell count, mean (SD), cells/mm^3	381.89 (221.28)
HCV genotype	
1a	19
1b	5
3a	4
4a	3
4d	6
HIV stage (SHIV)	
Non-C3 stage (NC3)	18
C3 stage (C3)	19
HIV load (LHIV)	
Low (L) <1000 copies/mL	29
High (H) >1000 copies/mL	8
Time since HCV infection (THCV), mean (SD), years	16.43 (0.97)
GGT level, mean (SD), U/L	104.3 (73.55)
ALT level, mean (SD), U/L	64.92 (55.91)
AST level, mean (SD), U/L	57.92 (33.74)
AST/ALT ratio (RAT), mean (SD)	1.00 (0.35)
TGC level, mean (SD), mg/dL	164.00 (118.50)
Platelet count (PLA), mean (SD), 10^9/L	194.57 (100.06)
Fibrosis indexes	
IF	
Mean (SD)	5.59 (1.74)
Median (range)	5.39 (1.99–11.65)
APRI	
Mean (SD)	1.02 (1.13)
Median (range)	0.77 (0.08–6.77)
FIB-4	
Mean (SD)	2.08 (2.46)
Median (range)	1.43 (0.32–15.36)

GGT, gamma-glutamyltransferase; ALT, alanine transaminase; AST, aspartate transaminase; TGC, triglycerides.

3.3. Statistical modelling

We applied a GLM framework (Venables and Ripley, 1999) to model the degree of liver fibrosis (response variable) according to several demographic, epidemiological, medical, and laboratory characteristics. The variables finally selected as potential predictors of fibrosis are shown in Tables 1 and 2 and Appendix A. We selected these variables based on the following criteria: (i), we chose the variables which have been widely identified as the most useful to detect liver damage, (ii) some variables included in the calculations of the Forns, APRI or FIB4 indexes were not introduced in the model (i.e. platelet count, cholesterol, AST, ALT), in order to avoid information duplication and (iii) variables with missing values for any patients in the cohort were not introduced in the model (for example, the HCV viral load that was absent in some patients).

In order to include the genetic variability as a variable in the statistical analyses we assigned each clade of the phylogenetic tree to a new variable. This new variable is then considered a binary predictor, with the value 1 or 0 if the clade is included or not, respectively (Szmaragd et al., 2006). Tables 1 and 2 show the mean and standard error of the predictor variables used in the analyses, separated by genotypes. Table 3 shows the correlation analyses between the three fibrosis indexes. There was a significant and high correlation among the three indirect indexes.

The variability parameters for HCV core nucleotide sequences are shown in Table 4. The *Haplotype diversity* is a measure of the uniqueness of a particular haplotype (sequence) in a given sample. The *Nucleotide diversity* is the average number of nucleotide differences *per site* between any two DNA sequences chosen randomly from the sample, and the *Nucleotide differences* are the mean number of different nucleotides among the sequences of a sample. In the 3a, 4a and 4d genotypes, the haplotype diversity was 1, indicating that all the sequences are different in each genotype: for this reason we found nucleotide differences in all genotypes. The highest genetic variability was found in genotype 3a isolates. To include HCV core variability in the GLM modelling, two separate GLM analyses were then performed, either using the 50% or the 75% majority rule in the phylogenetic tree. The HCV genotype was excluded as variable from the analyses, because HCV subtypes were coincident with clades (a–e), and therefore redundant with the inclusion of the clades as variables in the model. The statistical iterations were made to find the lowest *AIC* in each combination.

Table 2
The same variables indicated in Table 1 but separated by genotypes.

Variables	1a ($n = 19$)		1b ($n = 5$)		3a ($n = 4$)		4a ($n = 3$)		4d ($n = 6$)	
SEX	14 M	5 F	5 M		3 M	1 F	3 M		4 M	2 F
MofT	19 IDU		5 IDU		3 IDU	1 HET	3 IDU		4 IDU	2 HET
ENO	5 Y	14 N	1 Y	4 N	1 Y	3 N	1 Y	2 N	6 N	
HAART	17 Y	2 N	4 Y	1 N	4 Y		3 Y		6 Y	
SHIV	10 NC3	9 C3	2 NC3	3 C3	3NC3	1 C3	2 NC3	1 C3	1NC3	5 C3
LHIV	14 L	5 H	3 L	2 H	3 L	1 H	3 L		6 L	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
AGE	39.26	7.01	37.60	3.97	40.25	3.30	41.33	10.02	42.33	2.94
CD4	382.63	220.66	415.00	126.70	372.50	284.09	317.67	48.95	390.33	333.65
THCV	15.89	5.46	18.40	5.59	13.75	8.46	17.00	8.19	18.00	5.76
GGT	116.95	68.63	63.20	33.28	90.25	62.88	142.33	113.04	88.83	100.10
ALT	56.53	23.46	41.00	18.52	89.5	23.27	44.67	26.65	54.81	32.65
AST	54.95	21.66	51.20	29.22	104.00	78.55	54.00	34.77	62.0	51.91
RAT	1.09	0.40	0.88	0.23	1.05	0.39	0.86	0.22	0.85	0.22
TGC	200.00	144.48	132.20	86.79	60.75	32.99	124.67	45.88	165.00	55.17
PLA	184.89	59.58	268.00	210.57	122.50	88.69	177.33	8.02	184.86	99.04
Indexes										
IF	5.55	1.30	4.42	1.72	7.88	2.76	5.82	1.63	5.04	1.38
APRI	0.87	0.55	0.56	0.36	2.90	2.61	0.64	0.41	0.77	0.77
FIB-4	1.84	1.08	1.15	0.71	5.87	6.43	1.44	0.77	1.43	0.67

N = no; Y = yes; M = male; F = female; n = number of patients. Abbreviations as indicated in Table 1.

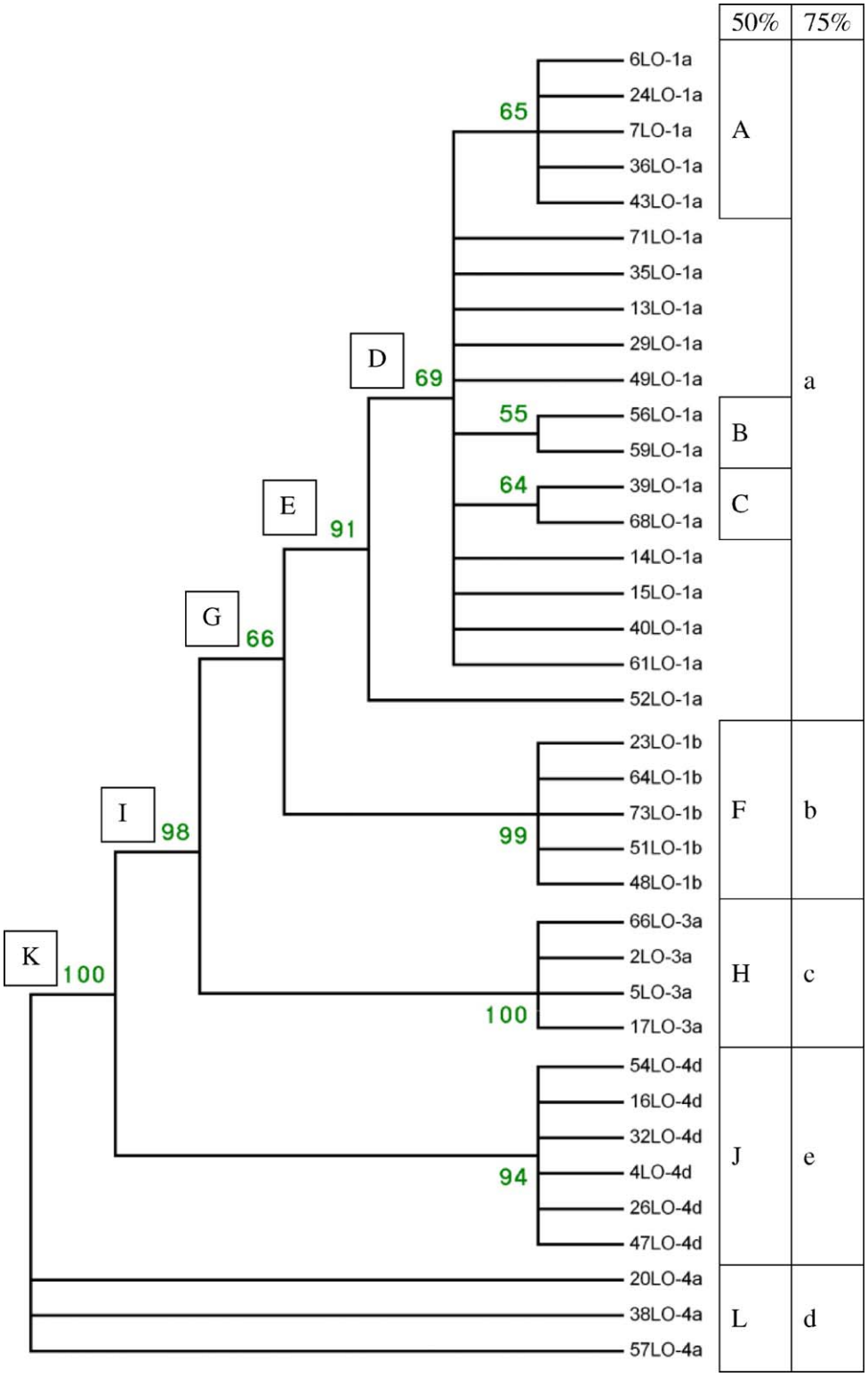

Fig. 1. The maximum parsimony consensus tree for the core region, supported by bootstrap (2000 replicates). Letters A–L represent clades with >50% support, and letters a–e represent clades with >75% support.

The fibrosis indexes data followed a normal distribution, and we therefore fitted normal distribution (family Gaussian with identity link). Table 5 shows the best models for the 50% and 75% majority rule. The GLM selected the best combination of predictors with the lowest *AIC* value for each response variable. Using the 50% majority rule, in the IF model analysis, statistically significant positive predictors were the AST/ALT ratio, the AGE and the antiretroviral treatment (HAART). The low HIV viral (LVIHL) and the serum triglycerides amount (TGC) showed a negative association. The phylogenetic cluster H (gt. 3a) from the consensus tree was positively significant in this analysis. The best fitted model, supported by the lowest *AIC*, in the APRI index, included as predictors, AST/ALT ratio (with a positive estimate), the IDU risk factor for HIV and HCV infection (MoftIDU) (with a negative and

Table 3
Correlation between the three fibrosis indexes.

	IF	APRI	FIB-4
IF	1.00		
APRI	0.80	1.00	
FIB-4	0.81	0.97	1.00

Table 4
Genetic variability parameters of the genotypes.

	N	Haplotype diversity	Nucleotide diversity	Nucleotide differences
Total	37	0.997	0.076	30.028
1a	19	0.989	0.014	5.373
1b	5	0.972	0.020	7.806
3a	4	1.000	0.040	15.900
4a	3	1.000	0.014	5.500
4d	6	1.000	0.009	3.733

statistically low significance estimate), and also the phylogenetic cluster H (gt. 3a). In the FIB-4 index, the model included the same predictors that in APRI, plus the AGE and cluster J (gt. 4d), both of them with low signification. In summary, variables positively correlated with fibrosis using the three indexes were the AST/ALT ratio and the phylogenetic cluster H. MoftIDU was the variable negatively correlated using APRI and FIB-4. The multivariate normal analysis (with the three response variables together) chose those predictor variables that better correlate with the three response variables at the same time. As can be seen, the most significant predictors in the model were AST/ALT ratio and the cluster H (gt. 3a), both positives. The AGE and MoftIDU contributions were statistically very low. The 75% majority bootstrap models are identical to those at the 50%, for this reason both analyses are unified in the same Table 5. In this case, the phylogenetic cluster was c (gt. 3a), equivalent to the cluster H at the 50%, and also e (gt. 4d), equivalent to J. In this Table, the 75% clusters selected are indicated in parentheses.

Table 6 shows the GLM analyses (with the three response variable together), but without the 17LO3a patient, that shows the highest values in the three fibrosis indexes (IF = 11.65, APRI = 6.77 and FIB-4 = 15.36). The two analyses are different between them (although they present coincidences in some predictor variables) and at the same time, different from the previous ones (when the 17LO3a patient is included). More variables were selected and the *AIC* was also lower. The AST/ALT ratio was positively included as was the MoftIDU (this variable was always negative in the previous analyses). Other new variables appear in these analyses: CD4, HAART, LHIV, SHIV or ENO. Also, another interesting aspect is the selected clusters. At 50%, are selected the A (negative, some patients with gt. 1a) and F (negative, the gt. 1b). At the 75%, the cluster selected was b (negative, gt. 1b). In both cases different from the cluster H (gt. 3a) selected in the previous analyses.

4. Discussion

We carried out an estimation of factors associated with liver fibrosis using the GLM methodology in a similar way as in Szmaragd et al. (2006) for HBV chronic infection, with the aim of identifying host clinical variables and HCV genetic features potentially linked to

Table 5
GLM analyses with 50% and 75% majority rule data. In parentheses it is indicated the cluster selected with 75%.

Coefficients	Estimate	Std. error	*t* value	Pr(>|*t*|)
IF glm (formula = IF ~ RAT + AGE + HAART + LHIV + TGC + H (c), family = gaussian, data = c50) (dispersion parameter for gaussian family taken to be 1.51) **AIC: 128.38**				
(Intercept)	−1.439	1.786	−0.805	0.427
RAT	2.523	0.608	4.148	0.0003***
AGE	0.109	0.038	2.858	0.008**
HAARTY	1.868	0.792	2.357	0.025*
LHIVL	−1.307	0.536	−2.437	0.021*
TGC	−0.004	0.002	−2.307	0.028*
H (c)	1.665	0.702	2.372	0.024*
Null deviance: 112.53 on 36 df, residual deviance: 45.16 on 30 df				
APRI glm (formula = APRI ~ RAT + Moft + H (c), family = gaussian, data = c50) (dispersion parameter for gaussian family taken to be 0.68) **AIC: 96.46**				
(Intercept)	0.383	0.536	0.715	0.480
RAT	1.217	0.397	3.067	0.004**
MoftIDU	−0.887	0.448	−1.980	0.056.
H (c)	1.895	0.444	4.270	0.0002***
Null deviance: 46.47 on 36 df, residual deviance: 22.41 on 33 df				
FIB-4 glm (formula = FIB-4 ~ RAT + AGE + Moft + H (c) + J (e), family = gaussian, data = c50) (dispersion parameter for gaussian family taken to be 2.58) **AIC: 147.58**				
(Intercept)	−1.411	2.348	−0.601	0.552
RAT	3.182	0.801	3.973	0.0004***
AGE	0.090	0.046	1.940	0.062.
MoftIDU	−3.694	1.073	−3.442	0.002**
H (c)	3.082	0.905	3.406	0.002**
J (e)	−1.873	0.927	−2.020	0.052.
Null deviance: 218.94 on 36 df, residual deviance: 80.09 on 31 df				
IF + APRI + FIB-4 glm (formula = IF + APRI + FIB-4 ~ RAT + AGE + Moft + H (c), family = gaussian, data = c50) (dispersion parameter for gaussian family taken to be 13.49) **AIC: 207.90**				
(Intercept)	−2.446	5.200	−0.470	0.641
RAT	6.906	1.810	3.816	0.0006***
AGE	0.154	0.105	1.475	0.150
MoftIDU	−3.162	1.998	−1.583	0.123
H (c)	7.977	1.979	4.031	0.0003***
Null deviance: 943.53 on 36 df, residual deviance: 431.66 on 32 df				

df: degrees of freedom, number of Fisher scoring iterations: 2,
.: $p<0.1$.
*: $p<0.05$.
**: $p<0.01$.
***: $p<0.001$.

Table 6
GLM analyses with 50%and 75% majority rule data with the three predictors together without the patient 17LO3a.

Coefficients	Estimate	Std. error	*t* value	Pr(>|*t*|)
IF + APRI + FIB-4 glm (formula = IF + APRI + FIB-4 ~ RAT + Moft + HAART + SEX + CD4 + SHIV + TGC + ENO + LHIV + A + F, family = gaussian, data = c50N) (dispersion parameter for gaussian family taken to be 1.96) **AIC: 137.87**				
(Intercept)	0.453	1.776	0.255	0.801
RAT	4.268	0.819	5.214	2.42e−05***
MoftIDU	3.940	1.078	3.655	0.001**
HAARTY	3.006	1.140	2.638	0.014*
SEXM	1.717	0.733	2.343	0.028*
CD4	0.003	0.001	2.068	0.050*
SHIVNC3	−1.621	0.636	−2.551	0.018*
TGC	−0.005	0.002	−2.089	0.047*
ENOY	−2.458	0.723	−3.401	0.002**
LHIVL	−3.558	0.714	−4.984	4.33e−05***
A	−2.772	0.880	−3.151	0.004**
F	−3.470	0.837	−4.144	0.0004***
Null deviance: 290.28 on 35 df, residual deviance: 47.14 on 24 df				
IF + APRI + FIB-4 glm (formula = IF + APRI + FIB-4 ~ RAT + Moft + HAART + CD4 + LHIV + ENO + b, family = gaussian, data = c75N) (Dispersion parameter for gaussian family taken to be 2.74) **AIC: 147.43**				
(Intercept)	−1.654	1.818	−0.909	0.371
RAT	5.048	0.927	5.443	8.28e−06***
MoftIDU	2.171	1.094	1.984	0.057.
HAARTY	5.206	1.133	4.595	8.39e−05***
CD4	0.004	0.002	2.648	0.013*
LHIVL	−3.763	0.828	−4.545	9.60e−05***
ENOY	−3.089	0.824	−3.750	0.0008***
b	−1.823	0.835	−2.183	0.038*
Null deviance: 290.28 on 35 df, residual deviance: 76.77 on 28 df				

df: degrees of freedom, number of Fisher scoring iterations: 2,
.: $p<0.1$.
*: $p<0.05$.
**: $p<0.01$.
***: $p<0.001$.

unfavourable outcomes of HCV disease in HCV–HIV coinfected individuals. This methodology allows the combination of phylogenetic tools with statistical modelling, thus combining the analysis of demographic, epidemiological and clinical features of the patients jointly with genetic data of the viral genome. The resulting models allow in turn the reduction of the number of predictor variables in the final analyses, and the choice of the model which best describes the response variable, in our case liver fibrosis.

As discussed above, one of the ways to detect liver damage is to evaluate the level of fibrosis, by means of a liver biopsy; but this is not always possible or available in retrospective studies from anonymous cohorts, such as that presented here. To overcome this limitation, we used three widely used indirect indexes: Forns, APRI and FIB-4 (Wai et al., 2003) to estimate the level of liver damage (Forns et al., 2002; Sterling et al., 2006; Vallet-Pichard et al., 2007). It its well known that these indexes may be non-accurate, although they seem a little more accurate for HCV–HIV coinfected patients (Ramos Paesa et al., 2007). In fact, we found this lack of precision when comparing the three indexes with data from fibrosis by biopsy in some patients (data not shown). The average correlation was only 30%, giving an idea of the low accuracy of the three indices to detect liver damage. This result is not only due to the low accuracy, but also to the low number of fibrosis data from biopsies, only 9 of the 37 patient. However, the three indexes showed high correlation between them (around the 86% on average), and this correlation was finally reflected in the GLM analyses, not only when they were made separately but also simultaneously. Therefore, the results showed here need to be interpreted with caution. Nevertheless, some authors have raised the question of whether liver biopsies have real diagnostic value for the evaluation of fibrosis, because of sampling error (Bedossa et al., 2003), although, nowadays, the detection of the fibrosis by biopsy is still considered as the Gold Standard. Besides, the invasiveness of the procedure and the possible severe complications makes desirable finding other surrogate variables evaluating fibrosis.

We used a maximum parsimony approach in the construction of phylogeny tree of HCV core nucleotide sequences. This approach has been previously shown to be more robust than maximum-likelihood when heterogeneity in mutation rates exist (Kolaczkowski and Thornton, 2004). As expected, all the HCV subtypes were grouped together when the 75% bootstrap majority rule was used in the analysis, but we detected significant clades within subtypes at nucleotide level in the 50% majority rule analyses, indicating some differential genetic variation within subtypes.

The genotype 3a isolates (“H” cluster in the 50% rule and “c” in the 75%) showed the highest genetic variability, and also the highest standard deviation in the fibrosis index. This is due to the presence of an infected patient (17LO3a) with the lowest platelet number. We decided not to exclude the patient from the study in order to observe the consequences of this potential outlier.

The GLM analyses were performed separately and also with the three indexes together (looking into a multivariate normal analysis), and following the methodology developed by Szmaragd et al. (2006) but with some modifications. The analyses were Gaussian instead of multinomial, because the three fibrosis indexes represented a broad range of values. For each model, some host variables and phylogenetic HCV clades were identified as predictors of fibrosis in our HCV–HIV coinfected cohort.

What do the GLM analyses results show us from a biological point of view? Some common features can be extracted from the different GLM analyses performed. In all the analyses, AST/ALT ratios are positively correlated with the fibrosis indexes, which indicate the importance of the transaminases indirectly revealing liver cell necrosis (Gressner et al., 2007). Statistically, it is possible to think that this correlation is a consequence of the incorporation of AST and/or ALT in the formula of APRI and FIB4. But nor AST or ALT appear in the IF formula, and the AST/ALT ratio has a high signification in this analysis, too. The same kind of explanation could be applied to GGT, used in the IF formula, but, in this case, this variable was not selected by any of the analyses based in the IF

index. In the case of the AGE, its presence in some analyses seems to be a consequence of the statistical process and not for the inclusion in some indexes.

The AST/ALT ratios are selected in all the analyses, positive and with high signification, probably because the patient with the highest fibrosis indexes (17LO3a) presents one of the highest value for this parameter (1.33). Observing the patient characteristics, the negative correlation of the MoftIDU with fibrosis seems striking (because he was not contagiated by IDU), although the estimates were lower than the AST/ALT ratios. It is possible that the patient 17LO3a could have also influenced the selection of the other variables in the model: high fibrosis indexes, relatively low triglycerides, HIV A3 stage (and so, NC3) and 45 years of age, being one of the oldest patients of the cohort. In fact, when the analyses are performed without this patient, the results are far different. The AST/ALT ratio continues being significant and following the same trend; but MoftIDU has changed the sign (now positive). At the same time, other variables become significants, as for example, the HAART regime, because most patients had been treated with antirretrovirals (34/37). These results suggest a significant effect of HAART in producing liver damage, as described by others (Matthews and Dore, 2008). The HIV viral load (LHIV) variable deserves an explanation, because apparently it is out of the trend found in the analysis with the other predictor variables. It is a variable that is selected by the IF index only, and with *negative* sign of the variant Low, when the 17LO3a patient presents a low value, and, so, it should be *positive*. A possible explanation could be that some patients that present also high fibrosis indexes have the variant H, influencing, in this way, the model. In addition, the IF index of the patient 17LO3a, presents lower deviation from the mean than the other two indexes. In the analysis without this patient, the LHIVL value is also negative and highly significant, confirming this explanation.

In all analyses with the 17LO3a patient included, the HCV cluster most selected was that related with gt. 3a (H in 50% bootstrap and c in 75%). Without him, the clusters selected are more diverse, this indicating the strong influence of this patient. More predictor variables are selected and in the clusters, have importance, mainly the genotypes 1a and 1b; more in accordance with the response of the patients to the liver damage.

5. Conclusions

This study shows that the GLM methodology can integrate different kind of host and viral genetic variables in one combined analysis, selecting those factors most likely to explain the response variable. The main finding of our study is the relation of some HCV clusters with the fibrosis biomarkers. The main limitations of our study are first, the use of indirect measures for the evaluation of liver fibrosis, because the retrospective and anonymous nature of the patient cohort and second, the lack of a control group of HCV monoinfected patients, to avoid confounders in evaluating the sole role of HIV coinfection in the development fibrosis. Other indirect noninvasive liver damage surrogate measures, such as liver elastography (Kotlyar et al., 2008) and/or data from liver biopsy evaluation will improve GLM estimates in further analyses. Some studies have associated the progression of the fibrosis with pre-disposition genes of the host (Sonzogni et al., 2002; Huang et al., 2006; Marcolongo et al., 2009). Thus, future studies using the GLM framework showed here may also include variables related with the genetic background of the infected patients and, ideally, genome-wide genetic polymorphism of both the virus and the host.

Acknowledgements

This work was supported by grant 4326/2007 from the Direcció General de Recerca, Desenvolupament Tecnològic i Innovació. Conselleria d'Economia, Hisenda i Innovació, Govern Balear, Spain and CIBEResp, Instituto de Salud Carlos III, Spain.

Appendix A. Parameters considered in the GLM analyses including the core region variability of the HCV (Arabic characters correspond to the clusters in the phylogenetic tree).

50%												75%																	Fibrosis indexes			Patients
A	B	C	D	E	F	G	H	I	J	K	L	a	b	c	d	e	GEN	LHIV	AGE	SEX	Moft	ENO	HAART	THCV	SHIV	CD4	RAT	GGT	TGC	IF	APRI	FIB-4	
1	0	0	1	1	0	1	0	1	0	1	0	1	0	0	0	0	1a	L	43	M	IDU	N	Y	22	C3	197	1.17	21	153	6.01	0.93	2.46	6LO1a
1	0	0	1	1	0	1	0	1	0	1	0	1	0	0	0	0	1a	L	48	M	IDU	N	Y	17	C3	300	1.13	22	153	4.94	0.48	1.67	24LO1a
1	0	0	1	1	0	1	0	1	0	1	0	1	0	0	0	0	1a	L	37	M	IDU	N	Y	21	C3	66	0.90	145	194	3.33	0.42	1.00	7LO1a
1	0	0	1	1	0	1	0	1	0	1	0	1	0	0	0	0	1a	H	39	M	IDU	N	N	21	NC3	468	1.06	147	76	5..87	0.58	1.33	36LO1a
1	0	0	1	1	0	1	0	1	0	1	0	1	0	0	0	0	1a	H	30	M	IDU	Y	Y	14	C3	340	1.06	166	200	5.48	1.11	1.48	43LO1a
0	0	0	1	1	0	1	0	1	0	1	0	1	0	0	0	0	1a	L	40	M	IDU	N	Y	18	NC3	403	0.89	93	240	5.85	1.12	2.21	71LO1a
0	0	0	1	1	0	1	0	1	0	1	0	1	0	0	0	0	1a	H	35	F	IDU	Y	N	2	NC3	813	1.05	81	110	5.34	0.20	0.63	35LO1a
0	0	0	1	1	0	1	0	1	0	1	0	1	0	0	0	0	1a	L	39	F	IDU	Y	Y	17	C3	520	1.84	128	177	5.40	1.31	2.85	13LO1a
0	0	0	1	1	0	1	0	1	0	1	0	1	0	0	0	0	1a	H	40	M	IDU	Y	Y	19	NC3	176	0.79	188	259	5.39	0.69	1.41	29LO1a
0	0	0	1	1	0	1	0	1	0	1	0	1	0	0	0	0	1a	L	44	M	IDU	N	Y	8	NC3	486	1.00	73	141	4.86	0.37	1.06	49LO1a
0	1	0	1	1	0	1	0	1	0	1	0	1	0	0	0	0	1a	L	35	F	IDU	N	Y	18	C3	316	0.88	134	264	4.80	0.87	1.41	56LO1a
0	1	0	1	1	0	1	0	1	0	1	0	1	0	0	0	0	1a	L	56	M	IDU	N	Y	15	C3	273	0.80	73	260	5.54	0.47	1.56	59LO1a
0	0	1	1	1	0	1	0	1	0	1	0	1	0	0	0	0	1a	H	35	M	IDU	N	Y	20	C3	4	2.06	176	276	8.40	2.50	4.90	39LO1a
0	0	1	1	1	0	1	0	1	0	1	0	1	0	0	0	0	1a	L	44	F	IDU	N	Y	15	C3	507	0.53	46	70	6.73	0.77	1.53	68LO1a
0	0	0	1	1	0	1	0	1	0	1	0	1	0	0	0	0	1a	L	30	F	IDU	N	Y	21	NC3	480	0.85	80	61	4.27	0.82	1.15	14LO1a
0	0	0	1	1	0	1	0	1	0	1	0	1	0	0	0	0	1a	L	32	M	IDU	N	Y	6	NC3	489	0.97	312	61	5.12	0.92	1.19	15LO1a
0	0	0	1	1	0	1	0	1	0	1	0	1	0	0	0	0	1a	L	38	M	IDU	Y	Y	18	NC3	844	1.06	91	148	6.29	0.75	1.93	40LO1a
0	0	0	1	1	0	1	0	1	0	1	0	1	0	0	0	0	1a	L	31	M	IDU	N	Y	12	NC3	147	1.84	87	242	8.32	1.83	4.07	61LO1a
0	0	0	0	1	0	1	0	1	0	1	0	1	0	0	0	0	1a	L	50	M	IDU	N	Y	18	NC3	441	0.74	159	715	3.6	0.5	1.1	52LO1a
0	0	0	0	0	1	1	0	1	0	1	0	0	1	0	0	0	1b	H	43	M	IDU	N	Y	22	C3	373	0.85	36	48	6.84	0.97	2.26	23LO1b
0	0	0	0	0	1	1	0	1	0	1	0	0	1	0	0	0	1b	L	40	M	IDU	N	N	24	NC3	326	1.24	100	270	1.99	0.08	0.32	64LO1b
0	0	0	0	0	1	1	0	1	0	1	0	0	1	0	0	0	1b	L	35	M	IDU	N	Y	14	C3	346	0.74	37	99	4.29	0.86	1.23	73LO1b
0	0	0	0	0	1	1	0	1	0	1	0	0	1	0	0	0	1b	L	37	M	IDU	N	Y	21	C3	637	0.94	44	85	4.74	0.41	1.05	51LO1b
0	0	0	0	0	1	1	0	1	0	1	0	0	1	0	0	0	1b	H	33	M	IDU	Y	Y	11	NC3	393	0.64	99	159	4.25	0.49	0.86	48LO1b
0	0	0	0	0	0	0	1	1	0	1	0	0	0	1	0	0	3a	L	38	M	IDU	N	Y	23	C3	400	1.40	121	51	7.09	2.10	4.01	66LO3a
0	0	0	0	0	0	0	1	1	0	1	0	0	0	1	0	0	3a	L	38	F	IDU	N	Y	12	NC3	763	0.55	41	21	5.06	1.24	1.26	2LO3a
0	0	0	0	0	0	0	1	1	0	1	0	0	0	1	0	0	3a	H	40	M	IDU	N	Y	17	NC3	136	0.94	164	99	7.74	1.48	2.85	5LO3a

Appendix A (*Continued*)

50%												75%																Fibrosis indexes			Patients
A	B	C	D	E	F	G	H	I	J	K	L	a	b	c	d	e	GEN	LHIV	AGE	SEX	Moft	ENO	HAART	THCV	SHIV	CD4	RAT	GGT	TGC	IF	APRI	FIB-4	
0	0	0	0	0	0	0	1	1	0	1	0	0	0	1	0	0	3a	L	45	M	HET	Y	Y	3	NC3	191	1.33	35	72	11.65	6.77	15.36	**17LO3a**
0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	1	4d	L	40	M	HET	N	Y	10	C3	396	0.63	49	201	3.71	0.27	0.73	54LO4d
0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	1	4d	L	46	M	IDU	N	Y	25	NC3	734	1.00	83	207	5.93	0.52	1.43	16LO4d
0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	1	4d	L	40	M	HET	N	Y	12	C3	58	0.78	45	233	2.96	0.22	0.53	32LO4d
0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	1	4d	L	42	M	IDU	N	Y	20	C3	837	0.59	289	114	6.03	2.29	2.13	4LO4d
0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	1	4d	L	46	F	IDU	N	Y	20	C3	56	1.15	20	103	5.33	0.53	1.67	26LO4d
0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	1	4d	L	40	M	HET	N	Y	10	C3	396	0.63	49	201	3.71	0.27	0.73	47LO4d
0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0	4a	L	42	M	IDU	N	Y	15	NC3	310	0.68	139	156	5.25	0.35	0.97	20LO4a
0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0	4a	L	31	M	IDU	Y	Y	10	NC3	273	1.10	31	72	4.54	0.46	1.02	38LO4a
0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0	4a	L	51	M	IDU	N	Y	26	C3	370	0.80	257	146	7.66	1.11	2.33	57LO4a

References

1993 revised classification system for HIV infection and expanded surveillance case definition for AIDS among adolescents and adults. MMWR Recomm. Rep. 41, 1–19.

Alter, M.J., 2006. Epidemiology of viral hepatitis and HIV co-infection. J. Hepatol. 44, S6–S9.

Alter, M.J., 2007. Epidemiology of hepatitis C virus infection. World J. Gastroenterol. 13, 2436–2441.

Bataller, R., North, K.E., Brenner, D.A., 2003. Genetic polymorphisms and the progression of liver fibrosis: a critical appraisal. Hepatology 37, 493–503.

Bedossa, P., Dargere, D., Paradis, V., 2003. Sampling variability of liver fibrosis in chronic hepatitis C. Hepatology 38, 1449–1457.

Bruguera, M., Forns, X., 2006. Hepatitis C in Spain. Med. Clin. 127, 113–117.

Cifuentes, C., Juan, M., Sastre, M., Payeras, A., Fernadez-Baca, V., Homar, F., Gallegos, C., Bassa, A., 2004. Coinfected by HIV–HCV: epidemiological data. First genotype determination in a sample from Mallorca Island.In: 27th World congress of Internal Medicine, Granada, Spain.

Crawley, M.J., 2007. The R book. John Wiley & Sons, Ltd., West Sussex.

Echevarría, J.M., León, P., Pozo, F., Avellón, A., 2006. Follow-up of the prevalence of hepatitis C virus genotypes in Spain during a nine-year period (1996–2004). Enferm. Infecc. Microbiol. Clin. 24, 20–25.

Esteban, J.I., Sauleda, S., Quer, J., 2008. The changing epidemiology of hepatitis C virus infection in Europe. J. Hepatol. 48, 148–162.

Forns, X., Ampurdanes, S., Llovet, J.M., Aponte, J., Quinto, L., Martinez-Bauer, E., Bruguera, M., Sanchez-Tapias, J.M., Rodes, J., 2002. Identification of chronic hepatitis C patients without hepatic fibrosis by a simple predictive model. Hepatology 36, 986–992.

Franco, S., Tural, C., Clotet, B., Martínez, M.A., 2007. Complete nucleotide sequence of genotype 4 hepatitis C viruses isolated from patients co-infected with human immunodeficiency virus type 1. Virus Res. 123, 161–169.

Gigou, M., Roque-Afonso, A.M., Falissard, B., Penin, F., Dussaix, E., Feray, C., 2001. Genetic clustering of hepatitis C virus strains and severity of recurrent hepatitis after liver transplantation. J. Virol. 75, 11292–11297.

Gressner, O.A., Weiskirchen, R., Gressner, A.M., 2007. Biomarkers of hepatic fibrosis, fibrogenesis and genetic pre-disposition pending between fiction and reality. J. Cell. Mol. Med. 11, 1031–1051.

Hall, T.A., 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucl. Acids Symp. Ser. 41, 95–98.

Heathcote, E.J., 2007. Antiviral therapy: chronic hepatitis C. J. Viral Hepat. 14 (Suppl 1), 82–88.

Huang, H., Shiffman, M.L., Cheung, R.C., Layden, T.J., Friedman, S., Abar, O.T., Yee, L., Chokkalingam, A.P., Schrodi, S.J., Chan, J., Catanese, J.J., Leong, D.U., Ross, D., Hu, X., Monto, A., McAllister, L.B., Broder, S., White, T., Sninsky, J.J., Wright, T.L., 2006. Identification of two gene variants associated with risk of advanced fibrosis in patients with chronic hepatitis C. Gastroenterology 130, 1679–1687.

Jimenez-Hernandez, N., Torres-Puente, M., Bracho, M.A., Garcia-Robles, I., Ortega, E., del Olmo, J., Carnicer, F., Gonzalez-Candelas, F., Moya, A., 2007. Epidemic dynamics of two coexisting hepatitis C virus subtypes. J. Gen. Virol. 88, 123–133.

Kolaczkowski, B., Thornton, J.W., 2004. Performance of maximum parsimony and likelihood phylogenetics when evolution is heterogeneous. Nature 431, 980–984.

Kotlyar, D.S., Blonski, W., Rustgi, V.K., 2008. Noninvasive monitoring of hepatitis C fibrosis progression. Clin. Liver Dis. 12, 557–571 viii.

Kuiken, C., Hraber, P., Thurmond, J., Yusim, K., 2008. The hepatitis C sequence database in Los Alamos. Nucl. Acids Res. 36, D512–D516.

Librado, P., Rozas, J., 2009. DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. Bioinformatics.

Lindenbach, B.D., Rice, C.M., 2005. Unravelling hepatitis C virus replication from genome to function. Nature 436, 933–938.

Lopez-Labrador, F.X., Ampurdanes, S., Forns, X., Castells, A., Saiz, J.C., Costa, J., Bruix, J., Sanchez Tapias, J.M., Jimenez de Anta, M.T., Rodes, J., 1997. Hepatitis C virus (HCV) genotypes in Spanish patients with HCV infection: relationship between HCV genotype 1b, cirrhosis and hepatocellular carcinoma. J. Hepatol. 27, 959–965.

Lopez-Labrador, F.X., Berenguer, M., Sempere, A., Prieto, M., Sirera, R., Gonzalez-Molina, A., Ortiz, V., Marty, M.L., Berenguer, J., Gobernado, M., 2004. Genetic variability of hepatitis C virus NS3 protein in human leukocyte antigen-A2 liver transplant recipients with recurrent hepatitis C. Liver Transpl. 10, 217–227.

Lopez-Labrador, F.X., Bracho, M.A., Berenguer, M., Coscolla, M., Rayon, J.M., Prieto, M., Carrasco, D., Gomez, M.D., Moya, A., Gonzalez-Candelas, F., 2006. Genetic similarity of hepatitis C virus and fibrosis progression in chronic and recurrent infection after liver transplantation. J. Viral Hepat. 13, 104–115.

Lopez-Labrador, F.X., Dove, L., Hui, C.K., Phung, Y., Kim, M., Berenguer, M., Wright, T.L., 2007. Trends for genetic variation of Hepatitis C Virus quasispecies in Human Immunodeficiency virus-1 coinfected patients. Virus Res. 130, 285–291.

Macias, J., Berenguer, J., Japon, M.A., Giron, J.A., Rivero, A., Lopez-Cortes, L.F., Moreno, A., Gonzalez-Serrano, M., Iribarren, J.A., Ortega, E., Miralles, P., Mira, J.A., Pineda, J.A., 2009. Fast fibrosis progression between repeated liver biopsies in patients coinfected with human immunodeficiency virus/hepatitis C virus. Hepatology 50, 1056–1063.

Marcolongo, M., Young, B., Dal Pero, F., Fattovich, G., Peraro, L., Guido, M., Sebastiani, G., Palu, G., Alberti, A., 2009. A seven-gene signature (cirrhosis risk score) predicts liver fibrosis progression in patients with initially mild chronic hepatitis C. Hepatology 50, 1038–1044.

Matthews, G.V., Dore, G.J., 2008. HIV and hepatitis C coinfection. J. Gastroenterol. Hepatol. 23, 1000–1008.

McCaughan, G.W., George, J., 2004. Fibrosis progression in chronic hepatitis C virus infection. Gut 53, 318–321.

Nagayama, K., Kurosaki, M., Enomoto, N., Miyasaka, Y., Marumo, F., Sato, C., 2000. Characteristics of hepatitis C viral genome associated with disease progression. Hepatology 31, 745–750.

Ohno, O., Mizokami, M., Wu, R.R., Saleh, M.G., Ohba, K., Orito, E., Mukaide, M., Williams, R., Lau, J.Y., 1997. New hepatitis C virus (HCV) genotyping system that allows for identification of HCV genotypes 1a, 1b, 2a, 2b, 3a, 3b, 4, 5a, and 6a. J. Clin. Microbiol. 35, 201–207.

Pybus, O.G., Cochrane, A., Holmes, E.C., Simmonds, P., 2005. The hepatitis C virus epidemic among injecting drug users. Infect. Genet. Evol. 5, 131–139.

Ramos Paesa, C., Marcilla, F., Lopez, G., Hueso, E., Pascual, A., Aguirre, J.M., 2007. Valuation of APRI and Forns models for non-invasive diagnosis of fibrosis in patients with hepatitis C in coinfected and non-coinfected with HIV. Med. Interna. 24, 369–374.

Reiberger, T., Ferlitsch, A., Sieghart, W., Kreil, A., Breitenecker, F., Rieger, A., Schmied, B., Gangl, A., Peck-Radosavljevic, M., 2010. HIV-HCV co-infected patients with low CD4+ cell nadirs are at risk for faster fibrosis progression and portal hypertension. J. Viral Hepat. 17, 400–409.

Rotman, Y., Liang, T.J., 2009. Coinfection with hepatitis C virus and human immunodeficiency virus: virological, immunological, and clinical outcomes. J. Virol. 83, 7366–7374.

Shuhart, M.C., Sullivan, D.G., Bekele, K., Harrington, R.D., Kitahata, M.M., Mathisen, T.L., Thomassen, L.V., Emerson, S.S., Gretch, D.R., 2006. HIV infection and antiretroviral therapy: effect on hepatitis C virus quasispecies variability. J. Infect. Dis. 193, 1211–1218.

Simmonds, P., Bukh, J., Combet, C., Deleage, G., Enomoto, N., Feinstone, S., Halfon, P., Inchauspe, G., Kuiken, C., Maertens, G., Mizokami, M., Murphy, D.G., Okamoto, H., Pawlotsky, J.M., Penin, F., Sablon, E., Shin-I, T., Stuyver, L.J., Thiel, H.J., Viazov, S., Weiner, A.J., Widell, A., 2005. Consensus proposals for a unified system of nomenclature of hepatitis C virus genotypes. Hepatology 42, 962–973.

Sonzogni, L., Silvestri, L., De Silvestri, A., Gritti, C., Foti, L., Zavaglia, C., Bottelli, R., Mondelli, M.U., Civardi, E., Silini, E.M., 2002. Polymorphisms of microsomal epoxide hydrolase gene and severity of HCV-related liver disease. Hepatology 36, 195–201.

Sterling, R.K., Lissen, E., Clumeck, N., Sola, R., Correa, M.C., Montaner, J.S., Sulkowski, M., Torriani, F.J., Dieterich, D.T., Thomas, D.L., Messinger, D., Nelson, M., 2006. Development of a simple noninvasive index to predict significant fibrosis in patients with HIV/HCV coinfection. Hepatology 43, 1317–1325.

Szmaragd, C., Nichols, R.A., Balloux, F., 2006. A novel approach to characterise pathogen candidate genetic polymorphisms involved in clinical outcome. Infect. Genet. Evol. 6, 38–45.

Tamura, K., Dudley, J., Nei, M., Kumar, S., 2007. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Mol. Biol. Evol. 24, 1596–1599.
Thomson, B.J., Finch, R.G., 2005. Hepatitis C virus infection. Clin. Microbiol. Infect. 11, 86–94.
Vallet-Pichard, A., Mallet, V., Nalpas, B., Verkarre, V., Nalpas, A., Dhalluin-Venier, V., Fontaine, H., Pol, S., 2007. FIB-4: an inexpensive and accurate marker of fibrosis in HCV infection. Comparison with liver biopsy and fibrotest. Hepatology 46, 32–36.
Venables, W.N., Ripley, B.D., 1999. Generalized Linear Models. Modern Applied Statistics with S. Springer, New York, pp. 183–210.
Wai, C.T., Greenson, J.K., Fontana, R.J., Kalbfleisch, J.D., Marrero, J.A., Conjeevaram, H.S., Lok, A.S., 2003. A simple noninvasive index can predict both significant fibrosis and cirrhosis in patients with chronic hepatitis C. Hepatology 38, 518–526.
Walters, K.A., Smith, M.W., Pal, S., Thompson, J.C., Thomas, M.J., Yeh, M.M., Thomas, D.L., Fitzgibbon, M., Proll, S., Fausto, N., Gretch, D.R., Carithers Jr., R.L., Shuhart, M.C., Katze, M.G., 2006. Identification of a specific gene expression pattern associated with HCV-induced pathogenesis in HCV- and HCV/HIV-infected individuals. Virology 350, 453–464.