Identification of a duplicated V3 domain in NS5A associated with cirrhosis and hepatocellular carcinoma in HCV-1b patients

H. Le Guillou-Guillemette, A. Ducancelle, S. Bertrais, C. Lemaire, A. Pivert, P. Veillon, E. Bouthry, S. Alain, Valérie Thibault, F. Abravanel, et al.

TITLE PAGE

Title: Identification of a duplicated V3 domain in NS5A associated with cirrhosis and hepatocellular carcinoma in HCV-1b patients

Authors: H Le Guillou-Guillemette [a,b], A Ducancelle [a,b], S Bertrais [b], C Lemaire [c], A Pivert [a,b], P Veillon [a,b], E Bouthry [a,b], S Alain [d], V Thibault [e], F Abravanel [f], AR Rosenberg [g], C Henquell [h], E André-Garnier [i], O Petsaris [j], S Vallet [j], JB Bour [k], Y Baazia [l], P Trimoulet [m], P André [n], C Gaudy-Graffin [o], D Bettinger [p], S Larrat [q], A Signori-Schmuck [q], H Saoudin [r], B Pozzetto [r], G Lagathu [s], S Minjolle-Cha [s], F Stoll-Keller [t], JM Pawlotsky [u], J Izopet [f], C Payan [j] and F Lunel-Fabiani [a,b].

[a] Laboratoire de Virologie, CHU Angers, France (heleguillou@chu-angers.fr; adpivert@chu-angers.fr; sandrine.bertrais@univ-angers.fr; paveillon@chu-angers.fr; alducancelle@chu-angers.fr; elise.bouthry@abc.aphp.fr; frlunel-fabiani@chu-angers.fr)

[b]HIFIH Laboratory, UPRES 3859, SFR 4208, LUNAM University, Angers, France

[c]IRHS, PRES LUNAM, SFR QUASAV, Angers, France (christophe.lemaire@univ-angers.fr)

[d] Laboratoire de Virologie, CHU Limoges, France (sophie.alain@unilim.fr)

[e]Laboratoire de Virologie, CHU Pitié-Salpêtrière, Paris, France (vincent.thibault@psl.aphp.fr)

[f]Laboratoire de Virologie, CNR VHE, Inserm U1043, CHU Purpan, Toulouse, France (abravanel.f@chu-toulouse.fr; izopet.j@chu-toulouse.fr)

[g]AP-HP, GHU Cochin, Laboratoire de Virologie, Université Paris Descartes, EA 4474 "Hepatitis C Virology", Paris, France (arielle.rosenberg@parisdescartes.fr)

[h]Laboratoire de Virologie, CHU Clermont-Ferrand, France (chenquell@chu-clermontferrand.fr)

[i] Laboratoire de Virologie, CHU Hôtel Dieu, Nantes, France (elisabeth.andre@chu-nantes.fr)

[j]Département de Bactériologie-Virologie-Hygiène Hospitalière et Parasitologie-Mycologie, CHRU, LUBEM, EA3882, Brest, France (odilepetsaris@yahoo.fr; sophie.vallet@chu-brest.fr; christopher.payan@chu-brest.fr)

[k] Laboratoire de Virologie, CHU, Dijon, France (Jean-Baptiste.Bour@u-bourgogne.fr)

[l] Laboratoire de Virologie, CHU Avicenne, Bobigny, France (yazid.baazia@sls.aphp.fr)

[m]Laboratoire de Virologie, Hôpital Pellegrin Tripode, CHU Bordeaux, France (pascale.trimoulet@chu-bordeaux.fr)

[n]Laboratoire de Virologie, Centre de Biologie Nord, Hôpital de la Croix Rousse, Lyon, France (patrice.andre@chu-lyon.fr)

[o]Université François Rabelais, Inserm U966, CHU Tours, France (gaudy_c@med.univ-tours.fr)

[p] Laboratoire de Virologie, CHU Besançon, France (dbettinger@chu-besançon.fr)

[q]Laboratoire de Virologie, UMI 3265 UJF-EMBL-CNRS, CHU, Unit of Virus Host Cell Interactions, Grenoble, France (SLarrat@chu-grenoble.fr; asignorischmuck@chu-grenoble.fr)

[r]Laboratoire de Bactériologie-Virologie, CHU Saint-Etienne, France (henia.saoudin@chu-st-etienne.fr; Bruno.Pozzetto@chu-st-etienne.fr)

[s]Laboratoire de Virologie, CHU Rennes, France (Gisele.LAGATHU@chu-rennes.fr; sophie.minjolle@chu-rennes.fr)

[t]Institut de Virologie, CHU Strasbourg, Inserm U748, Strasbourg, France (francoise.stoll@unistra.fr)

[u]Laboratoire de Virologie-Bactériologie, CHU Henri-Mondor, Créteil, France (jean-michel.pawlotsky@hmn.aphp.fr)

Corresponding author: Dr Hélène Le Guillou-Guillemette, Laboratoire de Virologie, Département des agents infectieux, IBS, CHU Angers, 4 rue Larrey 49933 Angers cedex9, France. Phone: 33-241354709. Fax: 33-241354164. E-mail: heleguillou@chu-angers.fr

Word counts: 2497 words

Abstract

Background. The NS5A protein of the hepatitis C virus has been shown to be involved in the development of hepatocellular carcinoma.

Objectives. In a French multicenter study, we investigated the clinical and epidemiological features of a new HCV genotype 1b strain bearing a wide insertion into the V3 domain.

Study Design. We studied NS5A gene sequences in 821 French patients infected with genotype 1b HCV.

Results. We identified an uncharacterized V3 insertion without ORF disruption in 3.05% of the HCV sequences. The insertion comprised 31 amino-acids for the majority of patients; 3 patients had 27 amino-acids insertions and 1 had a 12 amino-acids insertion. Sequence identity between the 31 amino-acids insertions and the V3 domain ranged from 48 to 96% with E-values above $4e^{-5}$, thus illustrating sequence homology and a partial gene duplication event that to our knowledge has never been reported in HCV. Moreover we showed the

presence of the duplication at the time of infection and its persistence at least during 12 years in the entire quasispecies. No association was found with extrahepatic diseases. Conversely, patients with cirrhosis were two times more likely to have HCV with this genetic characteristic (p=0.04). Moreover, its prevalence increased with liver disease severity (from 3.0% in patients without cirrhosis to 9.4% in patients with both cirrhosis and HCC, p for trend=0.045).

Conclusions. We identified a duplicated V3 domain in the HCV-1b NS5A protein for the first time. The duplication may be associated with unfavorable evolution of liver disease including a possible involvement in liver carcinogenesis.

Key words: cirrhosis, hepatitis C virus, hepatocellular carcinoma, NS5A, V3 domain

Background

Infection by hepatitis C virus (HCV) becomes persistent in around 80% of cases and in turn a major cause of liver disease for the approximately 184 million HCV patients worldwide. Chronically infected patients are at risk of developing liver cirrhosis and hepatocellular carcinoma (HCC) [1].

HCV is a single-strand RNA virus belonging to the *Flaviviridae* family. Its genome encodes a polyprotein of about 3000 amino-acids that is then cleaved into structural and non-structural proteins such as NS5A, a pleiotropic phosphoprotein. Thus, studies strongly suggest that NS5A plays a role in resistance to interferon (IFN)-based therapy [2], in the establishment of chronic hepatitis and in liver carcinogenesis [3]. NS5A is now a target for antivirals [4].

HCV is notable for its considerable genetic diversity [5]. Indels are one feature of the genetic variability characterizing RNA virus evolution, but the physiopathological consequences of these changes in HCV have not been well explored. Thus, Gerotto et al. showed that a 1-amino-acid insertion or deletion in HVR1 was significantly associated with mixed cryoglobulinemia type 2 [6]. Other authors have correlated indels in the NS5A domain with genotype specific signatures and suggested a putative role in the resistance to antiviral therapy or host immune response evasion [7]. Conversely, gene duplications have been described as a major evolutionary tool for DNA viruses [8]. This allows for the acquisition of new biological functions or the modification of virulence. However, biological constraints make gene duplication extremely rare in RNA viruses [9].

Objectives

In previous studies [10, 11] in HCV-1b-infected patients, we discovered a NS5A polymorphism that may consist in V3 domain duplication, a variable region located in the NS5A domain III. One of those patients had extrahepatic disease. To better understand this unknown but intriguing HCV polymorphism at NS5A gene, we performed an observational cross-sectional multicenter study. Thus, using both clinical and phylogenetic data, we questioned *i)* the impact of the duplication of V3 domain on the probability of HCC, *ii)* the prevalence of V3-duplication in patients and *iii)* the persistence of V3-duplicated strains during the infection process.

Study design

Study Patients

All consecutive patients newly diagnosed with chronic HCV-1b infection and without HIV co-infection were selected at the routine genotyping step in 19 French academic laboratory of virology between 2006 and 2009. We initially enrolled 938 patients; 117 patients were subsequently excluded due to failed NS5A amplification (HCV viral load was weak or undetectable). Excluded patients did not significantly differ regarding disease duration, date of diagnosis and alcohol abuse (data not shown). Consequently, the present cross-sectional study was conducted in 821 patients.

Demographic and clinical data

Date of blood sample collection and biological parameters including HCV genotype, viral load and biochemical measures were obtained from the participating laboratories. Demographic (age, gender) and clinical (fibrosis evaluation and presence of cirrhosis, presence of HCC, cryoglobulinemia, lymphoma or co-infection, presumed date of infection and route of transmission if available) data at the time of inclusion were retrospectively collected from medical records. Thus, we were able to determine fibrosis stages or the presence of cirrhosis in 542 patients.

NS5A amplification and nucleotide sequencing

HCV RNA was extracted from 200 µL of each serum sample using the EasyMag automated extraction system (BioMérieux, Craponne, France). As previously described [12], the full length NS5A gene was amplified by nested RT-PCR. Each PCR product that showed a larger

fragment than expected was sequenced in two directions using the Big Dye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems) on the automated ABI3130xl [12]. We also always used the Superscript III reverse transcriptase/platinum Taq polymerase (Invitrogen) to rule out PCR artifacts.

PCR products from consecutive serums with V3-insertion were also submitted to a clonal quasispecies analysis using the protocol that we previously described [10]. A mean of 30 clones per sample was analyzed. The amino-acids residues were numbered according to the sequence of the HCV-1b HCV-J prototype (Genbank Accession number: D90208).

Sequence analysis of the NS5A gene

When both were present, the amino-acids sequences of the inserted domain and the V3 domain were compared pairwise using BLASTP [13]. Sequences were considered homologous when the expected (E)-value was above 10^{-5}. This cut-off has already been used to assess duplication events in RNA viruses and was considered relatively stringent [14].

The sequence data were deposited in the DDBJ/EMBL/GenBank nucleotide sequence databases: accession numbers KF420489-KF420513.

Evolution of the V3-duplicated strains in one patient

In order to test for the functionality of strains carrying the duplication of V3 domain during infection, we performed the detection of the V3-insertion in 5 consecutive serums covering a twelve years follow-up of one patient. We had the serum before the contamination, at the time of the transmission in 1988 (by a kidney graft) and during the HCV chronic disease (1996, 1998 and 2000). In order to trace the evolutionary relationships of strains during time, we used the phylogenetic method implemented in the software BEAST v.1.8 [15]. Prior to any

inference the more likely substitution model was tested using jModelTest 2 [45]. The best model was found to be the HKY substitution model [46] with gamma heterogeneity and a proportion of invariant sites. This substitution model was used when useful in analyses using BEAST. Coalescent simulations assuming a strict clock model of substitution rate were run under the Bayesian Skyline model [16] that assumes demographic variations. Three independent runs were performed to assess convergence of the Monte Carlo Markov Chain process.

Statistical analyses

Statistical analyses were performed using SAS version 9.3 (SAS Institute, Inc., USA). The distribution of categorical variables was expressed as frequencies and percentages. Quantitative variables were described in mean values +/- standard deviation (SD) and median. Differences between patient groups were analyzed using the Fisher's exact test for nominal variables and the non-parametric Mann-Whitney-Wilcoxon test for continuous variables, unless otherwise indicated.

To assess the clinical relevance of the NS5A V3 insertion, we firstly compared the proportion of patients having this genetic characteristic as a function of several diseases. For this, 95% confidence intervals (CI) were calculated by the exact method assuming binomial distribution. We additionally performed an exact Cochran-Armitage trend test to assess whether the proportion of patients having the NS5A V3 insertion increased with liver disease severity (no cirrhosis, cirrhosis without HCC, cirrhosis with HCC). We also fitted multivariate ordinal logistic regression model to assess the relationship of the NS5A insertion with the 3 categories of liver disease severity. The proportional odds assumption underlying ordinal logistic regression was checked and was not rejected ($p>0.15$).Other factors which were

179 significantly related to liver disease severity in univariate analysis, were also tested in the
180 multivariate logistic regression model.

181

Results

Identification of a duplicated V3 domain in the NS5A gene

Twenty-five of the 821 test subjects were infected by an HCV-1b strain bearing a NS5A gene longer than expected. Direct NS5A sequencing of these samples revealed V3 domain duplication. Insertion lengths were 31 amino-acids for 21 patients, 27 amino-acids for three, and 12 amino-acids for one. The two V3 domains appeared in tandem without ORF disruption (Figure 2). The duplicated domains presented strong sequence similarities with a level of sequence identity ranging from 50 to 91%, and E-values ranged from $4e^{-5}$ to $2e^{-13}$.

The prevalence of this novel duplication in our cohort was 3.05% (95% CI: 1.98-4.46) without particular geographical localization.

Longitudinal analysis of the presence of the V3-insertion

A total of 112 different sequences of the NS5A gene were obtained across the four sampling times. All NS5A haplotypes carried the V3-insertion. The V3-insertion was then present since the infection and persisted for the 12 years. We did not observe any co-occurrence of haplotype of both haplotype with and without insertion within the same serum. The evolutionary relationship between haplotypes represented in Figure 2A showed the evolution of NS5A within the patient during the 12 years. Haplotype occurring at time t+1 derived from a small number of haplotypes present a time t. The bayesian skyline plot in Figure 2B indicates a quasi-constant population size during the 12 years. In addition, we also analysed consecutive serum available for two others patients covering a period of 5 and 6 years respectively. The V3-insertion was present in all the samples by direct and clonal sequencing.

Comparison between patients infected by the mutant HCV strain and those infected by wild-type HCV

Patient characteristics according to V3 status are shown in Table 1. The only difference between the two groups was that patients infected by an HCV virus bearing a V3 duplicated domain were older (p=0.035). No differences were found in terms of sex ratio, disease duration or HCV viral load.

In the subsample of patients for whom infection information was available, time since infection was highly related to age (Spearman rank correlation coefficient r=0.47, p<0.0001) but did not significantly differ according to V3 duplication status (p=0.91).

Clinical relevance of the V3 duplication

To assess possible relationships between the V3 duplication and extrahepatic disease, we classified patients according to the presence of cryoglobulinemia or lymphoma. No significant difference in the prevalence of V3 duplication was observed between patients having cryoglobulinemia or lymphoma as compared to those free of these extrahepatic diseases (Figure 3A).

In contrast, a significant relationship was found between V3 duplication and liver disease. The proportion of patients having the mutant strain was two times higher in patients with hepatocellular carcinoma compared to those without. We also found a significant association between the presence of the V3 duplication and cirrhosis (Figure 3B). Furthermore, the prevalence of the mutant strain increased with liver disease severity (p for trend=0.0449) (Table 2). Results from ordinal logistic regression analysis confirmed that there was a borderline significant relationship between V3 duplication and liver disease severity: the

unadjusted odds of having cirrhosis without or with HCC versus being non-cirrhotic was 2.4 times greater in patients with V3 duplication (OR [95% CI]: 2.41 [1.00-5.81], p=0.0502). Age and sex were also associated with liver disease severity (p=<0.0001 and p=0.016 respectively), no significant relationship was found with alcohol consumption (p=0.60) and HCV viral load (p=0.35). After adjustment for age and sex, the odds-ratio remained higher than 2.0, although it was non-significant (OR [95% CI]: 2.06 [0.84-5.02], p=0.1123). Due to the proportional odds assumption of the ordinal logistic regression model, the presence of V3 duplication was associated with the same two-fold increase in odds for having HCC versus the combined category “cirrhotic but no HCC” or “non-cirrhotic”, after controlling for age and sex.

Discussion

In our multicenter study, we found 25 cases with a high sequence identity between both V3 domains, illustrating partial gene duplication in NS5A. In our test population, the prevalence of the strain was 3.05% (95% CI: 1.98-4.46). Authors recently reported that gene duplication is infrequent in RNA viruses: they identified only nine cases of gene duplication among 1198 RNA virus species analyzed using a BLAST approach [14]. Specifically in HCV, insertion or deletion is quite rare. A large quasispecies study identified only infrequent, very small (from 1 to 4 amino-acids) insertions in the E1-E2 or NS5A regions of HCV-1. This work demonstrated that insertions were detected in all the specimens comprising sequential samples but not always in all the clones [17]. Here we detect an insertion that was not only large but also a tandemly repeated domain. Moreover, an analysis performed at the quasispecies level showed that the V3 duplication was present in 100% of the clones of the 25 mutant HCVs (n=870 clones, with a mean of 30 clones per sample). We also identified the V3 duplication following HCV infection and its persistence in three patients. This temporal analysis permitted to conclude that the clones with V3-insertions had the ability to infect patients and persist during several years like clones without insertion. Others studies have described small indels in the 5'UTR [18] or in the E1-E2 region, especially in the HVR1 subdomain [19]. In our work, the insertion contained 31 amino-acids for almost all of the mutant HCVs (21/25). Furthermore, Moradpour et al. demonstrated that domain III of NS5A tolerates wide heterologous insertion [20]. Our findings demonstrate that insertions are also naturally tolerated as suggested *in vitro*. In contrast, past reports of genomic HCV polymorphisms have mainly described recombinant strains; none described duplicated domains [21]. Gene duplication in RNA viruses is possible, but it is a contradictory evolutionary pathway. Duplication leads to increased genome size, which may lead to new evolutionary

opportunities, or conversely the enhancement of defects in production or replication [22]. The few cases of partial gene duplication reported in RNA viruses were located in untranslated regions of *Flaviviridae* [23, 24]. Thus, a high mutation rate offers a better evolutionary path for RNA viruses. Therefore, a prevalence of 3.05% for our mutated strain exhibiting duplication in the ORF should be considered as high.

HCV infection severity has been linked to host parameters and virological characteristics. A number of reports have associated amino-acids variations in NS5A with issues in antiviral therapy [2, 10, 27] or the evolution of liver disease toward cancer [28, 29]. Recently, particular mutations in NS3 and NS5A were also showed to be more closely associated with the development of HCC [30]. Our results suggest that the identified polymorphism in NS5A may also be associated with a pejorative progression of liver disease. We observed that this V3 domain duplication was much more prevalent in patients with cirrhosis and HCC. Moreover, this prevalence increased with liver disease severity, and the relationship with liver complications remained strong after adjustment for sex and age, although non-significant. Clinical data for liver fibrosis were available for 551 patients with a cohort of 119 cirrhosis patients. These data were in accordance with the level of severe disease expected in the physiopathological history of HCV disease and sufficient to perform robust statistical analyses. We were unable to identify other factors that may potentially influence liver disease severity. In the present study, alcohol consumption and viral load level were not related to liver disease severity. Unfortunately, other well-established confounder factors were not available or only in a subsample of subjects. Date and route of transmission were not frequently recorded, as mostly observed in prospective studies [31]. Time since infection may be one of the potential confounding factors of the observed relationships with liver complications. However this factor was indirectly controlled by age adjustment in our study.

Indeed, we found a high correlation between time since infection and age among patients for whom this information was available. Moreover, in this subsample, time since infection lost its significant relation to liver complications after age adjustment (data not shown).

Although we feel that the prevalence of the duplication in NS5A should be considered as elevated in the setting of RNA viruses and especially in that of HCV, the small number of patients with V3 insertions in our study population may lead to a lack of statistical power. This may explain the borderline or non-significant p-values and the wide confidence intervals, despite the strong observed differences. For these same reasons, the non-significant relationship with extrahepatic disease (cryoglobulinemia and lymphoma) should be interpreted cautiously. This kind of correlation was previously identified with single nucleotide insertions located in the HVR1-envelope [6]. Contrary to E1-E2, NS5A does not appear to be genuinely involved in immune system stimulation or clonal selection of lymphocytes. The V3 domain, located within NS5A domain III, has been shown to be dispensable for HCV replication *in vitro* [32], and our results suggest that its duplication does not influence viral load *in vivo*. Indeed, we observed no differences in viral loads as a function of V3 status. Here, we identified in a clinical study a wide duplication event in the NS5A protein of a HCV-1b strain that can be transmitted and persists, and may be associated with a higher risk of liver complications. The results of this cross-sectional study should be confirmed in a larger cohort.

Conflicts of interest:

Funding: This study was funded by the ANRS (Agence de Nationale de Recherche sur le VIH et les hépatites, appel d'offre Mars 2006).

Competing interests: none declared

Ethical approval: An information letter was sent to all patients and the study was approved by the Ethics Committee of the University Hospital of Brest (Avis CPP Ouest 6-15112006). The study was also approved and registered by the national commission for information technology and civil liberties (CNIL, 907057-09032007).

Acknowledgements

The authors thank K. Erwin for proofreading the English manuscript.

The authors thank the AC11-HCV Study and AC33-Virology groups from the ANRS, and the following clinicians: Angers - I Fouchard-Hubert, P Calès; Avicenne - D Roulot; Besançon - V Di Martino; Bordeaux - V de Lédinghen; Brest - JB Nousbaum, F Tanné; Clermont-Ferrand - A Abergel; Dijon - A Minello, P Hillon; Grenoble - JP Zarski, V Leroy, MN Hilleret; Limoges - V Loustaud-Ratti; Lyon - V Zoulim; Nantes - J Gournay; Paris-Cochin - P Podevin; Paris-Pitié Salpétrière - P Lebray; Rennes - D Guyader; Saint-Etienne - M. Ecochard, X. Roblin; Strasbourg - M. Doffoël; Toulouse - S Métivier, K Barange; Tours - L D'Alteroche, Y Bacq.

The authors thank A Lefebvre, E Muel, Y Desille, I Renard, A Plages, A Rodallec and W Merrouche for their assistance in clinical data recording.

References

[1] Omland LH, Jepsen P, Krarup H, Schonning K, Lind B, Kromann-Andersen H, et al. Increased mortality among persons infected with hepatitis C virus. Clin Gastroenterol Hepatol. 2011;9:71-8. [PubMed: 20888437]

[2] Enomoto N, Sakuma I, Asahina Y, Kurosaki M, Murakami T, Yamamoto C, et al. Mutations in the nonstructural protein 5A gene and response to interferon in patients with chronic hepatitis C virus 1b infection. N Engl J Med. 1996;334:77-81. [PubMed: 8531962]

[3] Macdonald A, Harris M. Hepatitis C virus NS5A: tales of a promiscuous protein. J Gen Virol. 2004;85:2485-502. [PubMed: 15302943]

[4] Gao M, Nettles RE, Belema M, Snyder LB, Nguyen VN, Fridell RA, et al. Chemical genetics strategy identifies an HCV NS5A inhibitor with a potent clinical effect. Nature. 2010;465:96-100. [PubMed: 20410884]

[5] Simmonds P, Bukh J, Combet C, Deleage G, Enomoto N, Feinstone S, et al. Consensus proposals for a unified system of nomenclature of hepatitis C virus genotypes. Hepatology. 2005;42:962-73. [PubMed: 16149085]

[6] Gerotto M, Dal Pero F, Loffreda S, Bianchi FB, Alberti A, Lenzi M. A 385 insertion in the hypervariable region 1 of hepatitis C virus E2 envelope protein is found in some patients with mixed cryoglobulinemia type 2. Blood. 2001;98:2657-63. [PubMed: 11675335]

[7] Cuevas JM, Gonzalez M, Torres-Puente M, Jimenez-Hernandez N, Bracho MA, Garcia-Robles I, et al. The role of positive selection in hepatitis C virus. Infect Genet Evol. 2009;9:860-6. [PubMed: 19463971]

[8] Shackelton LA, Holmes EC. The evolution of large DNA viruses: combining genomic information of viruses and their hosts. Trends Microbiol. 2004;12:458-65. [PubMed: 15381195]

[9] Holmes EC. Error thresholds and the constraints to RNA virus evolution. Trends Microbiol. 2003;11:543-6. [PubMed: 14659685]

[10] Veillon P, Payan C, Le Guillou-Guillemette H, Gaudy C, Lunel F. Quasispecies evolution in NS5A region of hepatitis C virus genotype 1b during interferon or combined interferon-ribavirin therapy. World J Gastroenterol. 2007;13:1195-203. [PubMed: 17451199]

[11] Lunel-Fabiani F, Veillon P, Le Guillou-Guillemette H, Redureau J, Coeffic B, Subra JF, et al. Duplication of the V3 domain in Nonstructural 5A Protein related to Hepatitis C Virus extrahepatic manifestations. In: Abstracts, editor. 45th Interscience Conference on Antimicrobial Agents and Chemotherapy (ICAAC). Washington, DC, USA: American Society for Microbiology; 2005, p. 485.

[12] Veillon P, Payan C, Gaudy C, Goudeau A, Lunel F. [Mutation analysis of ISDR and V3 domains of hepatitis C virus NS5A region before interferon therapy with or without ribavirin]. Pathol Biol (Paris). 2004;52:505-10. [PubMed: 15531113]

[13] Altschul SF, Wootton JC, Gertz EM, Agarwala R, Morgulis A, Schaffer AA, et al. Protein database searches using compositionally adjusted substitution matrices. FEBS J. 2005;272:5101-9. [PubMed: 16218944]

[14] Simon-Loriere E, Holmes EC. Gene Duplication Is Infrequent in the Recent Evolutionary History of RNA Viruses. Mol Biol Evol. 2013. [PubMed: 23486612]

[15] Drummond AJ, Suchard MA, Xie D, Rambaut A. Bayesian phylogenetics with BEAUti and the BEAST 1.7. Mol Biol Evol. 2012;29:1969-73. [PubMed: 22367748]

[16] Drummond AJ, Rambaut A, Shapiro B, Pybus OG. Bayesian coalescent inference of past population dynamics from molecular sequences. Mol Biol Evol. 2005;22:1185-92. [PubMed: 15703244]

[17] Torres-Puente M, Cuevas JM, Jimenez-Hernandez N, Bracho MA, Garcia-Robles I, Carnicer F, et al. Contribution of insertions and deletions to the variability of hepatitis C virus populations. J Gen Virol. 2007;88:2198-203. [PubMed: 17622623]

[18] Revie D, Alberti MO, Braich RS, Bayles D, Prichard JG, Salahuddin SZ. Discovery of significant variants containing large deletions in the 5'UTR of human hepatitis C virus (HCV). Virol J. 2006;3:82. [PubMed: 17010198]

[19] Hohne M, Schreier E, Roggendorf M. Sequence variability in the env-coding region of hepatitis C virus isolated from patients infected during a single source outbreak. Arch Virol. 1994;137:25-34. [PubMed: 7979995]

[20] Moradpour D, Evans MJ, Gosert R, Yuan Z, Blum HE, Goff SP, et al. Insertion of green fluorescent protein into nonstructural protein 5A allows direct visualization of functional hepatitis C virus replication complexes. J Virol. 2004;78:7400-9. [PubMed: 15220413]

[21] Morel V, Fournier C, Francois C, Brochot E, Helle F, Duverlie G, et al. Genetic recombination of the hepatitis C virus: clinical implications. J Viral Hepat. 2011;18:77-83. [PubMed: 21235686]

[22] Pybus OG, Rambaut A, Belshaw R, Freckleton RP, Drummond AJ, Holmes EC. Phylogenetic evidence for deleterious mutation load in RNA viruses and its contribution to viral evolution. Mol Biol Evol. 2007;24:845-52. [PubMed: 17218639]

[23] Gritsun TS, Gould EA. The 3' untranslated region of tick-borne flaviviruses originated by the duplication of long repeat sequences within the open reading frame. Virology. 2006;354:217-23. [PubMed: 17063566]

[24] Gritsun TS, Gould EA. Direct repeats in the 3' untranslated regions of mosquito-borne flaviviruses: possible implications for virus transmission. J Gen Virol. 2006;87:3297-305. [PubMed: 17030864]

[25] Nagai M, Sakoda Y, Mori M, Hayashi M, Kida H, Akashi H. Insertion of cellular sequence and RNA recombination in the structural protein coding region of cytopathogenic bovine viral diarrhoea virus. J Gen Virol. 2003;84:447-52. [PubMed: 12560578]

[26] Trento A, Casas I, Calderon A, Garcia-Garcia ML, Calvo C, Perez-Brena P, et al. Ten years of global evolution of the human respiratory syncytial virus BA genotype with a 60-nucleotide duplication in the G protein gene. J Virol. 2010;84:7500-12. [PubMed: 20504933]

[27] Puig-Basagoiti F, Forns X, Furcic I, Ampurdanes S, Gimenez-Barcons M, Franco S, et al. Dynamics of hepatitis C virus NS5A quasispecies during interferon and ribavirin therapy in responder and non-responder patients with genotype 1b chronic hepatitis C. J Gen Virol. 2005;86:1067-75. [PubMed: 15784900]

[28] De Mitri MS, Morsica G, Cassini R, Bagaglio S, Zoli M, Alberti A, et al. Prevalence of wild-type in NS5A-PKR protein kinase binding domain in HCV-related hepatocellular carcinoma. J Hepatol. 2002;36:116-22. [PubMed: 11804673]

[29] Hung CH, Chen CH, Lee CM, Wu CM, Hu TH, Wang JH, et al. Association of amino acid variations in the NS5A and E2-PePHD region of hepatitis C virus 1b with hepatocellular carcinoma. J Viral Hepat. 2008;15:58-65. [PubMed: 18088246]

[30] El-Shamy A, Shindo M, Shoji I, Deng L, Okuno T, Hotta H. Polymorphisms of the core, NS3 and NS5A proteins of hepatitis C 1 virus genotype 1b associate with development of hepatocellular carcinoma. Hepatology. 2013;58:555-63. [PubMed: 23281009]

[31] Zeuzem S, Teuber G, Lee JH, Ruster B, Roth WK. Risk factors for the transmission of hepatitis C. J Hepatol. 1996;24:3-10. [PubMed: 8836883]

[32] Tellinghuisen TL, Foss KL, Treadaway JC, Rice CM. Identification of residues required for RNA replication in domains II and III of the hepatitis C virus NS5A protein. J Virol. 2008;82:1073-83. [PubMed: 18032500]

418
419 **FIGURE CAPTIONS**

420

421 **Figure 1. Alignments of the 25 sequences showing a V3-NS5A duplicated domain with**
422 **respect to the HCV-J 1b reference sequence.** The first V3 domain is located between
423 positions 2353 and 2383 (light grey column), the second is tandemly located thereafter out to
424 position 2414 (dark grey column). Dots (.) represent identical residues; dashes (-) indicate
425 shorter insertions in the last four aligned sequences; X denotes a mixture.

426

427 **Figure 2. Phylodynamics at gene NS5A for 112 strains of a patient during its infection**
428 **since the year of the transmission (1988) until the last date of the clinical follow-up**
429 **(2000). (A)** Coalescent tree inferred using the Bayesian Skyline Model using BEASTv1.8.
430 Colors correspond to sampling times of quasispecies (1988: yellow; 1996: orange; 1998: red
431 and 2000: pink). **(B)** Demographic dynamics using the Bayesian Skyline Plot. Distribution of
432 effective population sizes are plotted against time. Black line corresponds to the median of
433 distribution and 95% of the highest posterior density is given in blue.

434

435 **Figure 3. Comparisons of the proportion of patients having the V3-duplicated HCV**
436 **strain according to clinical status for two extrahepatic diseases cryoglobulinemia and**
437 **lymphoma (A) and to the presence of HCC or cirrhosis (B).** Error bars represent 95%
438 confidence intervals.

439

Table 1. Characteristics of included patients according to HCV-1b NS5A V3 status

	Total	Presence of V3 duplication		
	n=821	No (n=796)	Yes (n=25)	
	n (%) or mean (sd) / median	n (%) or mean (sd) / median	n (%) or mean (sd) / median	p
Male (%)	50.3	50.3	48	0.84 [(a)]
Age (years)	54.8 (14.6) / 54.7	54.6 (14.6) / 54.3	61.0 (14.2) / 59.7	0.035[(b)]
Viral load (log UI/mL)	5.91 (0.78) / 5.98	5.91 (0.78) / 5.99	5.82 (0.62) / 5.88	0.43 [(b)]
Time since infection (years)	n=191 23.8 (11.9) / 23.0	n=176 23.8 (12.2) / 23.0	n=15 24.7 (9.3) / 22.0	0.91 [(b)]
Infection route known	n=279	n=265	n=14	
Transfusion	n=191 (68.5)	n=179 (67.6)	n=12 (85.7)	0.24 [(b)]
Injected drug use	n=43 (15.4)	n=42 (15.8)	n=1 (7.1)	0.70 [(b)]

[(a)] Fisher's exact test; [(b)] Non-parametric Mann–Whitney–Wilcoxon test

Table 2. Patient characteristics and observed prevalence of V3 duplication according to combined status for cirrhosis and hepatocellular carcinoma.

No cirrhosis	Cirrhosis without HCC	Cirrhosis with HCC

Variable	n=432 % or mean (sd) / median	n=87 % or mean (sd) / median	n=32 % or mean (sd) / median	p for linear trend
Male (%)	44.1	51.8	68.7	0.0056 [(a)]
Age (years)	54.1 (14.1) / 53.9	59.9 (12.2) / 62.0	65.2 (9.7) / 66.6	<0.0001 [(b)]
Viral load (log IU/ml)	5.91 (0.74) / 5.97	5.94 (0.79) / 6.00	5.67 (0.80) / 5.72	0.17 [(b)]
V3 duplication (%), 95% CI	3.01 (1.61-5.09)	5.75 (1.89-12.90)	9.38 (1.98-25.02)	**0.0449** [(a)]

[(a)] Cochran-Armitage test for trend; [(b)] General linear model including a linear contrast to test trend