



HAL
open science

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, Florence Legrand-Abravanel, et al.

► To cite this version:

H. Le Guillou-Guillemette, A. Ducancelle, S. Bertrais, C. Lemaire, A. Pivert, et al.. Identification of a duplicated V3 domain in NS5A associated with cirrhosis and hepatocellular carcinoma in HCV-1b patients. *Journal of Clinical Virology*, 2015, 69, pp.203-209. 10.1016/j.jcv.2015.06.096 . hal-01231431

HAL Id: hal-01231431

<https://univ-rennes.hal.science/hal-01231431v1>

Submitted on 25 Nov 2015

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

1

2 **TITLE PAGE**

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

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

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

14 ^bHIFIH Laboratory, UPRES 3859, SFR 4208, LUNAM University, Angers, France

15 ^cIRHS, PRES LUNAM, SFR QUASAV, Angers, France (christophe.lemaire@univ-angers.fr)

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

17 ^eLaboratoire de Virologie, CHU Pitié-Salpêtrière, Paris, France (vincent.thibault@psl.aphp.fr)

18 ^fLaboratoire de Virologie, CNR VHE, Inserm U1043, CHU Purpan, Toulouse, France
19 (abravanel.f@chu-toulouse.fr; izopet.j@chu-toulouse.fr)

20 ^gAP-HP, GHU Cochin, Laboratoire de Virologie, Université Paris Descartes, EA 4474
21 "Hepatitis C Virology", Paris, France (arielle.rosenberg@parisdescartes.fr)

22 ^hLaboratoire de Virologie, CHU Clermont-Ferrand, France (chenquell@chu-clermontferrand.fr)
23

- 24 ⁱ Laboratoire de Virologie, CHU Hôtel Dieu, Nantes, France (elisabeth.andre@chu-nantes.fr)
- 25 ^jDépartement de Bactériologie-Virologie-Hygiène Hospitalière et Parasitologie-Mycologie,
26 CHRU, LUBEM, EA3882, Brest, France (odilepetsaris@yahoo.fr; [sophie.vallet@chu-
brest.fr](mailto:sophie.vallet@chu-
27 brest.fr); christopher.payan@chu-brest.fr)
- 28 ^k Laboratoire de Virologie, CHU, Dijon, France (Jean-Baptiste.Bour@u-bourgogne.fr)
- 29 ^lLaboratoire de Virologie, CHU Avicenne, Bobigny, France (yazid.baazia@sls.aphp.fr)
- 30 ^mLaboratoire de Virologie, Hôpital Pellegrin Tripode, CHU Bordeaux, France
31 (pascale.trimoulet@chu-bordeaux.fr)
- 32 ⁿLaboratoire de Virologie, Centre de Biologie Nord, Hôpital de la Croix Rouse, Lyon, France
33 (patrice.andre@chu-lyon.fr)
- 34 ^oUniversité François Rabelais, Inserm U966, CHU Tours, France ([gaudy_c@med.univ-
tours.fr](mailto:gaudy_c@med.univ-
35 tours.fr))
- 36 ^p Laboratoire de Virologie, CHU Besançon, France (dbettinger@chu-besancon.fr)
- 37 ^qLaboratoire de Virologie, UMI 3265 UJF-EMBL-CNRS, CHU, Unit of Virus Host Cell
38 Interactions, Grenoble, France (SLarrat@chu-grenoble.fr; asignorischmuck@chu-grenoble.fr)
- 39 ^rLaboratoire de Bactériologie-Virologie, CHU Saint-Etienne, France ([henia.saoudin@chu-st-
etienne.fr](mailto:henia.saoudin@chu-st-
40 etienne.fr); [Bruno.Pozzetto@chu-st-etienne.fr](mailto: Bruno.Pozzetto@chu-st-etienne.fr))
- 41 ^sLaboratoire de Virologie, CHU Rennes, France (Gisele.LAGATHU@chu-rennes.fr;
42 sophie.minjolle@chu-rennes.fr)
- 43 ^tInstitut de Virologie, CHU Strasbourg, Inserm U748, Strasbourg, France
44 (francoise.stoll@unistra.fr)
- 45 ^uLaboratoire de Virologie-Bactériologie, CHU Henri-Mondor, Créteil, France ([jean-
michel.pawlotsky@hmn.aphp.fr](mailto:jean-
46 michel.pawlotsky@hmn.aphp.fr))

47

48 **Corresponding author:** Dr H el ene Le Guillou-Guillemette, Laboratoire de Virologie,
49 D epartement des agents infectieux, IBS, CHU Angers, 4 rue Larrey 49933 Angers cedex9,
50 France. Phone: 33-241354709. Fax: 33-241354164. E-mail: heleguillou@chu-angers.fr

51

52 **Word counts:** 2497 words

53

54

55

56 **Abstract**

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

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

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

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

69 presence of the duplication at the time of infection and its persistence at least during 12 years
70 in the entire quasispecies. No association was found with extrahepatic diseases. Conversely,
71 patients with cirrhosis were two times more likely to have HCV with this genetic
72 characteristic ($p=0.04$). Moreover, its prevalence increased with liver disease severity (from
73 3.0% in patients without cirrhosis to 9.4% in patients with both cirrhosis and HCC, p for
74 trend=0.045).

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

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

79

80 **Background**

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

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

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

100

101

102 **Objectives**

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

111

112 **Study design**

113 *Study Patients*

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

121

122 *Demographic and clinical data*

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

130 *NS5A amplification and nucleotide sequencing*

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

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

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

142 *Sequence analysis of the NS5A gene*

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

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

149 *Evolution of the V3-duplicated strains in one patient*

150 In order to test for the functionality of strains carrying the duplication of V3 domain during
151 infection, we performed the detection of the V3-insertion in 5 consecutive serums covering a
152 twelve years follow-up of one patient. We had the serum before the contamination, at the time
153 of the transmission in 1988 (by a kidney graft) and during the HCV chronic disease (1996,
154 1998 and 2000). In order to trace the evolutionary relationships of strains during time, we
155 used the phylogenetic method implemented in the software BEAST v.1.8 [15]. Prior to any

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

163 *Statistical analyses*

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

170 To assess the clinical relevance of the NS5A V3 insertion, we firstly compared the proportion
171 of patients having this genetic characteristic as a function of several diseases. For this, 95%
172 confidence intervals (CI) were calculated by the exact method assuming binomial distribution.
173 We additionally performed an exact Cochran-Armitage trend test to assess whether the
174 proportion of patients having the NS5A V3 insertion increased with liver disease severity (no
175 cirrhosis, cirrhosis without HCC, cirrhosis with HCC). We also fitted multivariate ordinal
176 logistic regression model to assess the relationship of the NS5A insertion with the 3
177 categories of liver disease severity. The proportional odds assumption underlying ordinal
178 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

Accepted Manuscript

182 **Results**

183 *Identification of a duplicated V3 domain in the NS5A gene*

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

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

192 *Longitudinal analysis of the presence of the V3-insertion*

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

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

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

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

212 *Clinical relevance of the V3 duplication*

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

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

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

235

236 Discussion

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

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

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

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

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

303

304 **Conflicts of interest:**

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

307 Competing interests: none declared

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

312

313 **Acknowledgements**

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

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

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

325

326 **References**

- 327 [1] Omland LH, Jepsen P, Krarup H, Schonning K, Lind B, Kromann-Andersen H, et al. Increased
328 mortality among persons infected with hepatitis C virus. *Clin Gastroenterol Hepatol*. 2011;9:71-8.
329 [PubMed: 20888437]
- 330 [2] Enomoto N, Sakuma I, Asahina Y, Kurosaki M, Murakami T, Yamamoto C, et al. Mutations in the
331 nonstructural protein 5A gene and response to interferon in patients with chronic hepatitis C virus 1b
332 infection. *N Engl J Med*. 1996;334:77-81. [PubMed: 8531962]
- 333 [3] Macdonald A, Harris M. Hepatitis C virus NS5A: tales of a promiscuous protein. *J Gen Virol*.
334 2004;85:2485-502. [PubMed: 15302943]
- 335 [4] Gao M, Nettles RE, Belema M, Snyder LB, Nguyen VN, Fridell RA, et al. Chemical genetics strategy
336 identifies an HCV NS5A inhibitor with a potent clinical effect. *Nature*. 2010;465:96-100. [PubMed:
337 20410884]
- 338 [5] Simmonds P, Bukh J, Combet C, Deleage G, Enomoto N, Feinstone S, et al. Consensus proposals
339 for a unified system of nomenclature of hepatitis C virus genotypes. *Hepatology*. 2005;42:962-73.
340 [PubMed: 16149085]
- 341 [6] Gerotto M, Dal Pero F, Loffreda S, Bianchi FB, Alberti A, Lenzi M. A 385 insertion in the
342 hypervariable region 1 of hepatitis C virus E2 envelope protein is found in some patients with mixed
343 cryoglobulinemia type 2. *Blood*. 2001;98:2657-63. [PubMed: 11675335]
- 344 [7] Cuevas JM, Gonzalez M, Torres-Puente M, Jimenez-Hernandez N, Bracho MA, Garcia-Robles I, et
345 al. The role of positive selection in hepatitis C virus. *Infect Genet Evol*. 2009;9:860-6. [PubMed:
346 19463971]
- 347 [8] Shackelton LA, Holmes EC. The evolution of large DNA viruses: combining genomic information of
348 viruses and their hosts. *Trends Microbiol*. 2004;12:458-65. [PubMed: 15381195]
- 349 [9] Holmes EC. Error thresholds and the constraints to RNA virus evolution. *Trends Microbiol*.
350 2003;11:543-6. [PubMed: 14659685]
- 351 [10] Veillon P, Payan C, Le Guillou-Guillemette H, Gaudy C, Lunel F. Quasispecies evolution in NS5A
352 region of hepatitis C virus genotype 1b during interferon or combined interferon-ribavirin therapy.
353 *World J Gastroenterol*. 2007;13:1195-203. [PubMed: 17451199]
- 354 [11] Lunel-Fabiani F, Veillon P, Le Guillou-Guillemette H, Redureau J, Coeffic B, Subra JF, et al.
355 Duplication of the V3 domain in Nonstructural 5A Protein related to Hepatitis C Virus extrahepatic
356 manifestations. In: Abstracts, editor. 45th Interscience Conference on Antimicrobial Agents and
357 Chemotherapy (ICAAC). Washington, DC, USA: American Society for Microbiology; 2005, p. 485.
- 358 [12] Veillon P, Payan C, Gaudy C, Goudeau A, Lunel F. [Mutation analysis of ISDR and V3 domains of
359 hepatitis C virus NS5A region before interferon therapy with or without ribavirin]. *Pathol Biol (Paris)*.
360 2004;52:505-10. [PubMed: 15531113]
- 361 [13] Altschul SF, Wootton JC, Gertz EM, Agarwala R, Morgulis A, Schaffer AA, et al. Protein database
362 searches using compositionally adjusted substitution matrices. *FEBS J*. 2005;272:5101-9. [PubMed:
363 16218944]
- 364 [14] Simon-Loriere E, Holmes EC. Gene Duplication Is Infrequent in the Recent Evolutionary History
365 of RNA Viruses. *Mol Biol Evol*. 2013. [PubMed: 23486612]
- 366 [15] Drummond AJ, Suchard MA, Xie D, Rambaut A. Bayesian phylogenetics with BEAUti and the
367 BEAST 1.7. *Mol Biol Evol*. 2012;29:1969-73. [PubMed: 22367748]
- 368 [16] Drummond AJ, Rambaut A, Shapiro B, Pybus OG. Bayesian coalescent inference of past
369 population dynamics from molecular sequences. *Mol Biol Evol*. 2005;22:1185-92. [PubMed:
370 15703244]

- 371 [17] Torres-Puente M, Cuevas JM, Jimenez-Hernandez N, Bracho MA, Garcia-Robles I, Carnicer F, et
372 al. Contribution of insertions and deletions to the variability of hepatitis C virus populations. *J Gen*
373 *Virol.* 2007;88:2198-203. [PubMed: 17622623]
- 374 [18] Revie D, Alberti MO, Braich RS, Bayles D, Prichard JG, Salahuddin SZ. Discovery of significant
375 variants containing large deletions in the 5'UTR of human hepatitis C virus (HCV). *Viol J.* 2006;3:82.
376 [PubMed: 17010198]
- 377 [19] Hohne M, Schreier E, Roggendorf M. Sequence variability in the env-coding region of hepatitis C
378 virus isolated from patients infected during a single source outbreak. *Arch Virol.* 1994;137:25-34.
379 [PubMed: 7979995]
- 380 [20] Moradpour D, Evans MJ, Gosert R, Yuan Z, Blum HE, Goff SP, et al. Insertion of green fluorescent
381 protein into nonstructural protein 5A allows direct visualization of functional hepatitis C virus
382 replication complexes. *J Virol.* 2004;78:7400-9. [PubMed: 15220413]
- 383 [21] Morel V, Fournier C, Francois C, Brochot E, Helle F, Duverlie G, et al. Genetic recombination of
384 the hepatitis C virus: clinical implications. *J Viral Hepat.* 2011;18:77-83. [PubMed: 21235686]
- 385 [22] Pybus OG, Rambaut A, Belshaw R, Freckleton RP, Drummond AJ, Holmes EC. Phylogenetic
386 evidence for deleterious mutation load in RNA viruses and its contribution to viral evolution. *Mol Biol*
387 *Evol.* 2007;24:845-52. [PubMed: 17218639]
- 388 [23] Gritsun TS, Gould EA. The 3' untranslated region of tick-borne flaviviruses originated by the
389 duplication of long repeat sequences within the open reading frame. *Virology.* 2006;354:217-23.
390 [PubMed: 17063566]
- 391 [24] Gritsun TS, Gould EA. Direct repeats in the 3' untranslated regions of mosquito-borne
392 flaviviruses: possible implications for virus transmission. *J Gen Virol.* 2006;87:3297-305. [PubMed:
393 17030864]
- 394 [25] Nagai M, Sakoda Y, Mori M, Hayashi M, Kida H, Akashi H. Insertion of cellular sequence and RNA
395 recombination in the structural protein coding region of cytopathogenic bovine viral diarrhoea virus.
396 *J Gen Virol.* 2003;84:447-52. [PubMed: 12560578]
- 397 [26] Trento A, Casas I, Calderon A, Garcia-Garcia ML, Calvo C, Perez-Brena P, et al. Ten years of global
398 evolution of the human respiratory syncytial virus BA genotype with a 60-nucleotide duplication in
399 the G protein gene. *J Virol.* 2010;84:7500-12. [PubMed: 20504933]
- 400 [27] Puig-Basagoiti F, Fornis X, Furcic I, Ampurdanes S, Gimenez-Barcons M, Franco S, et al. Dynamics
401 of hepatitis C virus NS5A quasispecies during interferon and ribavirin therapy in responder and non-
402 responder patients with genotype 1b chronic hepatitis C. *J Gen Virol.* 2005;86:1067-75. [PubMed:
403 15784900]
- 404 [28] De Mitri MS, Morsica G, Cassini R, Bagaglio S, Zoli M, Alberti A, et al. Prevalence of wild-type in
405 NS5A-PKR protein kinase binding domain in HCV-related hepatocellular carcinoma. *J Hepatol.*
406 2002;36:116-22. [PubMed: 11804673]
- 407 [29] Hung CH, Chen CH, Lee CM, Wu CM, Hu TH, Wang JH, et al. Association of amino acid variations
408 in the NS5A and E2-PePHD region of hepatitis C virus 1b with hepatocellular carcinoma. *J Viral Hepat.*
409 2008;15:58-65. [PubMed: 18088246]
- 410 [30] El-Shamy A, Shindo M, Shoji I, Deng L, Okuno T, Hotta H. Polymorphisms of the core, NS3 and
411 NS5A proteins of hepatitis C 1 virus genotype 1b associate with development of hepatocellular
412 carcinoma. *Hepatology.* 2013;58:555-63. [PubMed: 23281009]
- 413 [31] Zeuzem S, Teuber G, Lee JH, Ruster B, Roth WK. Risk factors for the transmission of hepatitis C. *J*
414 *Hepatol.* 1996;24:3-10. [PubMed: 8836883]
- 415 [32] Tellinghuisen TL, Foss KL, Treadaway JC, Rice CM. Identification of residues required for RNA
416 replication in domains II and III of the hepatitis C virus NS5A protein. *J Virol.* 2008;82:1073-83.
417 [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

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

	Total	Presence of V3 duplication		p
	n=821	No (n=796)	Yes (n=25)	
	n (%) or mean (sd) / median	n (%) or mean (sd) / median	n (%) or mean (sd) / median	
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)

441 ^(a) Fisher's exact test; ^(b) Non-parametric Mann-Whitney-Wilcoxon test

442

443

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

No cirrhosis	Cirrhosis without HCC	Cirrhosis with HCC
--------------	-----------------------	--------------------

Variable	n=432	n=87	n=32	p for linear trend
	% or mean (sd) / median	% or mean (sd) / median	% or mean (sd) / median	
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)

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

447

448

449

450

451