Title

Phylogenetic analysis of an epidemic outbreak of acute hepatitis C in HIV-infected patients by ultra-deep pyrosequencing.

Authors and Affiliations (First Name/Last Name)

Noelia Caro-Pérez[a], María Martínez-Rebollar[b], Josep Gregori[c,d], Josep Quer[c], Patricia González[a], Martina Gambato[a], Neris García-González[e], Fernando González-Candelas[e], Josep Costa[f], Juan Ignacio Esteban[c], Josep Mallolas[b], Xavier Forns[a], Montse Laguno[b]*, Sofía Pérez-del-Pulgar[a]*.

[a] Liver Unit, Hospital Clínic, IDIBAPS, CIBERehd, Universitat de Barcelona, Barcelona, Spain.

[b] Infectious Diseases Unit, Hospital Clínic, IDIBAPS, Barcelona, Spain.

[c] Liver Unit, Vall d'Hebron Institut de Recerca-Hospital Universitari Vall d'Hebron, CIBERehd, Barcelona, Spain.

[d] Roche Diagnostics SL. Sant Cugat del Vallès. Barcelona, Spain

[e] Joint Research Unit Infección y Salud Pública, FISABIO-Universitat de València, I2SysBio, CIBERESP, Valencia, Spain.

[f] Microbiology Department, Centre Diagnòstic Biomèdic, Hospital Clínic, CIBERehd, Barcelona, Spain.

*These authors contributed equally to this work and share senior authorship.

E-mail addresses

Noelia Caro-Pérez: caro@clinic.ub.es

María Martínez-Rebollar: rebollar@clinic.ub.es

Josep Gregori: josep.gregori@vhir.org

Josep Quer: josep.quer@vhir.org

Patricia González: pgonzale@clinic.ub.es

Martina Gambato: gambato@clinic.ub.es

Neris García-González: negargon@alumni.uv.es

Fernando González-Candelas: fernando.gonzalez@uv.es

Josep Costa: jcosta@clinic.ub.es

Juan Ignacio Esteban: jignacio.esteban@ciberehd.org

Josep Mallolas: mallolas@clinic.ub.es

Xavier Forns: xforns@clinic.ub.es

Montse Laguno: mlaguno@clinic.ub.es

Sofía Pérez-del-Pulgar: sofiapp@clinic.ub.es

Corresponding author

Sofía Pérez-del-Pulgar

CIBERehd, Hospital Clínic, IDIBAPS

Rossellón 149-153, 08036 Barcelona, Spain

Phone. +34 932275400 (2093) / Fax. +34 933129405

E-mail: sofiapp@clinic.ub.es

Word count

Abstract: 238 words

Main text: 2476 words

Highlights

- HCV infection has spread among HIV-infected MSM through a local network in Barcelona.

- A novel method based on low genetic differentiation is proposed to define transmission groups.

- Several potential clusters are identified: 8 for gt1a, 1 for gt1b and 7 for gt4d.

- HCV gt4d strains cluster with HIV-infected MSM from other European countries.

1 **Abstract**

2 Background: The incidence of acute hepatitis C (AHC) among HIV-infected men who

3 have sex with men (MSM) has increased significantly in the last 10 years. Several

4 studies point to a social and sexual network of HIV-positive MSM that extends

5 internationally.

6 Objectives: The aim of our study was to investigate the dynamics of HCV transmission

7 in an outbreak of AHC in HIV-infected MSM in Barcelona by ultra-deep pyrosequencing.

8 Study design: Between 2008 and 2013, 113 cases of AHC in HIV-infected MSM were

9 diagnosed in the Infectious Diseases Unit, Hospital Clínic, Barcelona. Massive

10 sequencing was performed using the Roche 454 GS Junior platform. To define possible

11 transmission networks, maximum likelihood phylogenetic trees were constructed, and

12 levels of genetic diversity within and among patients were compared.

13 Results: Among the 70 cases analyzed, we have identified 16 potential clusters of

14 transmission: 8 for genotype 1a (23 cases involved), 1 for genotype 1b (3 cases) and 7

15 for genotype 4d (27 cases). Although the initial phylogenetic reconstruction suggested

16 a local transmission cluster of HCV gt4d, our approach based on low genetic

17 differentiation did not corroborate it. Indeed, gt4d strains formed 4 independent

18 groups related to patients from other countries.

19 Conclusions: Frequent clustering of HIV-positive MSM shows that HCV infection has

20 spread through a local network in Barcelona. This outbreak is related to a large

21 international HCV transmission network among MSM. Public health efforts are needed

22 to reduce HCV transmission among this high-risk group.

23 **Keywords**

Abstract

26 AHC, acute hepatitis C; ALT, alanine aminotransferase; AST, aspartate

27 aminotransferase; gt, genotype; HCV, hepatitis C virus; HIV, human immunodeficiency

28 virus; IDU, intravenous drug users; ML, maximum likelihood; MSM, men who have sex

29 with men; nt, nucleotide.

30

1. Background and objectives

Hepatitis C virus (HCV) is mainly spread by contact with contaminated blood after transfusion or by sharing needles with patients who use drugs intravenously. In the past 10 years, a growing number of cases of acute hepatitis C (AHC) in HIV-positive men who have sex with men (MSM) have been reported in large urban centers in Europe [1–6], the United States [7], Asia [8] and Australia [9]. The increased incidence of AHC in HIV-positive MSM has been attributed to several factors: the tendency towards a higher HCV viral load in HIV patients' blood and semen [10], sexual practices with an increased risk of mucosal damage, the presence of ulcerative sexually transmitted diseases such as syphilis or lymphogranuloma venereum, a larger number of sexual partners, and the use of recreational drugs [11].

In 2009, a phylogenetic analysis of transmitted HCV strains showed clustering consistent with transmission in a social and sexual network of HIV-positive MSM that extends internationally [6]. HCV genotype (gt) 4 is the most prevalent genotype in the Middle East and in Northern and Central Africa, whereas in Europe, North America and Australia, HCV gt1 and gt3 are the most frequent ones [12]. Nonetheless, several authors have reported a high rate of HCV gt4d among HIV-positive MSM and identified monophyletic clusters, which suggests HCV spreading among MSM in Europe [5,13].

In this study, we have analyzed the transmission dynamics in an epidemic outbreak of AHC in HIV-positive MSM patients in Barcelona by ultra-deep pyrosequencing. The initial phylogenetic analyses suggested a single, local transmission cluster of HCV-gt4d. However, by using a novel approach that took intra- and inter-patient genetic diversities into account, we showed that this was not the case. Interestingly, this was

54 further confirmed because these gt4d HCV strains clustered in separate groups with an

55 international transmission network of HIV-positive MSM [6].

56

57 **2. Study design**

58 *2.1. Study population*

59 The Infectious Diseases Unit of the Hospital Clínic in Barcelona conducts routine

60 follow-up of HIV-infected patients every 6 months, including clinical, biochemical

61 evaluation and liver function tests, aspartate aminotransferase (AST) and alanine

62 aminotransferase (ALT). Anti-HCV antibody testing is performed at the time of

63 diagnosis of HIV infection, every two years, and whenever patients report risk factors

64 for HCV infection. All patients who have an unexplained increase in liver enzyme

65 activities (more than twice the upper normal limit) during routine monitoring of HIV

66 infection and all patients with clinical signs of acute hepatitis (jaundice, severe fatigue,

67 urine and stool discoloration) are screened for serum anti-HCV and HCV-RNA levels.

68 Between January 2008 and December 2013, 113 episodes of AHC in HIV-infected MSM

69 were diagnosed at our center [3]. Serum samples were available from 84 cases of AHC

70 that occurred in 79 patients. For comparison purposes, we included a control group

71 consisting of 29 chronic hepatitis C patients diagnosed in the area of Barcelona and

72 including all the genotypes involved in the AHC cases. The study was approved by the

73 Ethics Committee of the Hospital Clínic, Barcelona, and all the participants provided

74 written informed consent.

75

2.2. HCV-NS5B amplification for ultra-deep pyrosequencing

77 To analyze the dynamics of transmission in our cohort, a fragment of 340 bp of the

78 HCV-NS5B region (from nucleotides 8279 to 8618, isolate H77, accession AF009606)

79 was amplified and massively sequenced as previously described [14]. Data treatment

80 was performed using the pipelines designed to obtain clean nucleotide haplotypes [15]

81 and for HCV subtyping [14]. Full details of the PCR strategy and ultra-deep

82 pyrosequencing are provided as Supplementary Materials and Methods.

83 *2.3. HCV-NS5B amplification for direct sequencing*

84 In order to establish the relationships between the AHC cases reported here and an

85 international sexual transmission network of HIV-positive MSM [6], we amplified and

86 directly sequenced another fragment of 430 bp of the NS5B region (from nucleotides

87 8553 to 8982of the H77 reference sequence) as described elsewhere [16,17], but with

88 some modifications (additional details in Supplementary Materials and Methods).

89 *2.4. Phylogenetic analysis*

90 Multiple alignments were performed with ClustalW [18]. Maximum likelihood (ML)

91 trees were obtained with PhyML [19], using the GTR+Gamma substitution model as

92 determined by Modeltest v.3.8 [20]. Support for the internal nodes was evaluated with

93 1,000 bootstrap replicates [21]. Genetic diversity (intra- and inter-patient) was

94 calculated with Arlequin [22]. For the analysis of deep pyrosequencing data, we only

95 considered haplotypes with abundances above 0.5%.

96 To strengthen the identification of groups of transmission, we combined phylogenetic

97 clustering in the ML trees with the analysis of the net inter-patient genetic diversity

98 compared to the intra-patient genetic diversity in each group. We defined the net

99   inter-patient genetic diversity as the total genetic diversity of a group minus the

100   weighted average of the intra-patient diversities of the samples included in that group.

101   By using this approach, clusters in the phylogenetic tree must fulfill two conditions to

102   be considered as transmission groups: 1) high bootstrap support (≥80%) and 2) the net

103   inter-patient genetic diversity of their sequences (i.e. among the members of the

104   cluster) should be lower than the highest intra-patient diversity in the same genotype.

105   *2.5. Nucleotide sequence accession numbers*

106   NS5B sequences obtained in this study have been deposited in GenBank under

107   accession numbers SAMN06289805 to SAMN06289918 (deep sequencing, SRA

108   database BioProject PRJNA369605) and KY674814 to KY674856 (Sanger sequencing).

109

110   **3. Results**

111   *3.1. Baseline characteristics of patients*

112   Patient characteristics at the time of AHC diagnosis are shown in Table 1. The median

113   age was 39 years (range 34-44). In two cases, AHC and HIV diagnoses were

114   concomitant. Sixty-four patients were receiving combined antiretroviral therapy

115   before AHC diagnosis. At the time of AHC diagnosis, all but one patient had

116   undetectable HIV viral load, and median HCV-RNA was 6.28 Log IU/mL (range 4.78-

117   6.66). The prevalence of HCV genotypes was: 43 (51%) gt4d, 33 (39%) gt1a, 7 (8%) gt1b

118   and 1 (1%) gt3a (the latter was excluded from the ensuing analyses). Among the 84

119   AHC cases, 5 were reinfected with a different HCV genotype from that of the first

120   episode (Supplementary Table S1).

121   *3.2. Phylogenetic analysis of HCV quasispecies*

122   The HCV NS5B gene was successfully amplified and sequenced in 70 AHC cases (32

123   gt1a, 4 gt1b and 34 gt4d) and in the 29 local controls (8 gt1a, 7 gt1b and 14 gt4d).

124   Ultra-deep pyrosequencing yielded 225,698 reads from haplotypes with abundance

125   above 0.5%, with a median average coverage of 2,280 reads per sample

126   (Supplementary Table S2). Notably, no multiple HCV infections were found in any of

127   the cases analyzed.

128   To identify monophyletic clusters, we constructed ML trees for each genotype (Fig. 1).

129   Based on phylogenetic clustering (bootstrap support >80%), several putative

130   transmission clusters could be identified: 16 for gt1a (A to P, Fig. 1a), 2 for gt1b (A and

131   B, Fig. 1b) and 12 for gt4d (A to L, Fig. 1c). Interestingly, all gt4d AHC patients clustered

132   separately from the local controls in a well-supported clade. This was not the case for

133   gt1a and gt1b AHC patients, for which several clusters contained sequences from the

134   local controls. These observations suggested a single source of infection for gt4d in our

135   study cohort.

136   In order to assess whether the monophyletic groups identified by phylogenetic

137   clustering corresponded to actual transmission clusters, we considered that low net

138   genetic differentiation was an indicator of close epidemiological relatedness. First, we

139   computed the intra-patient genetic diversity for all the samples. The results indicated

140   that the local controls had the highest genetic diversity in all cases (Supplementary

141   Table S3). Next, we determined the net genetic diversity among AHC cases included in

142   the clusters identified in the ML tree (Supplementary Table S4). For gt1a, we identified

143   8 groups (B, C, F, G, H, I J and K, Fig. 1a) which had a net genetic diversity lower than

144 the maximum intra-patient value for this genotype (0.00944, control K3786). These

145 clusters included 5 patients at most. Cluster F was composed of subclusters D and E

146 and thus, they should be considered as a single transmission group. Likewise, only one

147 cluster of transmission was considered for gt1b, which included 3 of the 5 gt1b

148 analyzed patients (group A, Fig. 1b)

149 As mentioned before, all gt4d sequences clustered into a single clade with high

150 statistical support (group A, Fig. 1c). However, since its net divergence (0.01468) was

151 larger than the maximum intra-patient diversity for this subtype (0.01165, control

152 K2876), it could not be considered as a single cluster of transmission. A similar result

153 was obtained for group B, which included 16 patients, with a net diversity of 0.01537.

154 Thus, for gt4d, we observed 7 clusters (C, F, G, I, J, K and L, Fig. 1c) in which between 2

155 to 9 patients were involved.

156 *3.3. Comparison of genotype 4d with the international network*

157 Previous works have described a large, multinational network of HCV gt4d spreading

158 among MSM in several European countries [5,13]. To compare the gt4d sequences

159 from our study with those from the international network, we amplified and

160 sequenced another region of the NS5B as reported previously [6,16].

161 As shown in Fig 2, the ML tree derived from this international network of HCV

162 transmission and our local AHC and control sequences of gt4d clearly confirmed that

163 the AHC monophyletic group observed hitherto (group A, Fig. 1c) was no longer a

164 cluster of transmission. In particular, sequences from patients p45, p49, p51, p53, p57,

165 p62 and p64 clustered with high bootstrap support with sequences from Germany, the

166 Netherlands and France. These sequences also represented three independent

167 introductions: 1) p45 and p62; 2) p49, p51, p57 and p64; and 3) p53, in addition to the

168 group encompassing exclusively sequences from the remaining gt4d AHC patients. The

169 sequences from local controls were related only to one sequence obtained from the

170 international MSM network.

171

172 **4. Discussion**

173 As observed in several cities in Europe, North America and Australia, the incidence of

174 AHC among HIV-positive MSM has also increased exponentially in Spain during the last

175 10 years [2,3,23,24]. However, phylogenetic data in this group of patients are scarce

176 [25]. In this study, we have analyzed the network of transmission of an AHC epidemic

177 outbreak in HIV-infected MSM in Barcelona by ultra-deep pyrosequencing. In our

178 cohort, we identified 5 patients with one further AHC episode with a different HCV

179 genotype from that of the prior infection. Reinfection rate was 6%, somewhat less

180 than in previous studies, most probably due to shorter follow-ups, but still indicating

181 the lack of immune protection after HCV infection [26,27].

182 For epidemiologists, clusters represent an urgent need to intervene on focal groups to

183 prevent further spread and risk for the general population and, for these reasons, their

184 identification is relevant. However, from a molecular epidemiology point of view there

185 are many possible definitions of transmission cluster. Grabowski and Redd [28]

186 reviewed and summarized cluster definitions from 20 phylogenetic studies of HIV

187 transmission. Indeed, only 3 studies considered the same definition and that was the

188 least stringent of all: 70% bootstrap support in a ML or neighbor joining tree. In a

189 recent study based on consensus Sanger sequencing, Lamoury et al. showed that the

190   use of longer HCV sequences increased the accuracy of cluster identification [29].

191   Alternatively, deep sequencing of shorter fragments along with phylogenetic analyses

192   have also been applied in different settings to explore the relationships between

193   infected individuals [30–33]. In particular, NS5B deep sequencing has been shown to

194   improve the discrimination of clusters versus consensus Sanger sequencing analysis

195   when applying dynamic patristic distance thresholds based upon each individual's

196   intra-host viral population [33].

197   Here, we have introduced a new criterion to fine tune the identification of

198   transmission clusters after deep sequencing of a short fragment. We have compared

199   the net inter-patient genetic diversity among sequences from patients putatively

200   included in a cluster, because of strong grouping in a phylogenetic reconstruction (high

201   bootstrap in the ML tree), with the intra-patient diversity of a local control group of

202   chronic hepatitis C patients. We assumed that the largest intra-patient diversity

203   represented a reasonable upper limit to consider sequences with a lower net inter-

204   patient diversity among them as involved in a transmission chain. By applying this

205   criterion, we have identified 16 potential clusters of transmission: 8 for gt1a (23 cases

206   involved), 1 for gt1b (3 cases) and 7 for gt4d (27 cases).

207   Using intra-patient diversity from individuals with chronic HCV infection may be a

208   limitation, because the longer the chronic infection in a patient, the higher the

209   diversity of the viral population. Nonetheless, in our study cohort, the criterion of the

210   largest intra-patient diversity as an upper limit to consider closely related sequences

211   proved to be more accurate than those based on statistical support of the groupings.

212   For instance, both groups L and O in gt1a included various subclusters that fit well

213 within the adopted definition of transmission clusters because of their low net

214 diversity (subclusters J and K for group L, and subclusters B, C, D, E and F for group O).

215 However, groups L and O/P also included control patients such as K2084 in group L and

216 K0417 in groups O/P, which suggested that those groups might not be MSM

217 transmission clusters. According to this observation, the net divergences of groups L, O

218 and P were higher than the corresponding upper limit and thus, they were not

219 considered as transmission clusters.

220 Recent studies emphasize that gt4d is highly prevalent among intravenous drug users

221 (IDU) and HIV-positive MSM [5,6,13,16]. Indeed, the homogeneity of these strains

222 suggests that gt4d was introduced in the European IDU population quite recently,

223 entering the MSM population afterwards [11,16]. Our data show that 7 of our gt4d

224 AHC sequences grouped with sequences from this international network with at least 3

225 independent introductions, indicating that the outbreak in Barcelona is highly related

226 to previously described European MSM transmission clusters [6]. The existence of

227 epidemiological networks of HCV transmission among HIV-positive MSM has also been

228 confirmed in a recent study in which two clusters of AHC samples were highly related

229 to gt1a isolates from Germany, the Netherlands, United Kingdom and Australia,

230 belonging to the above mentioned transmission network [25].

231 One of the main limitations of the study is its retrospective nature, which limits the

232 epidemiological information, and has prevented us from identifying and dating the

233 source of the outbreak, particularly for gt4d. To collect data on social behavior and the

234 common leisure-oriented public places, self-administered questionnaires were given to

235 the patients. However, only half of them filled the forms completely and most

236 admitted being under the effect of recreational drugs, making it difficult to trace

237 common social habits.

238 In conclusion, our findings show that HCV infection has spread rapidly among HIV-

239 infected MSM through a local network in Barcelona. Frequent clustering of gt4d AHC

240 patients with other European HIV-coinfected cases demonstrates that this outbreak is

241 related to an interconnected transmission network of transnational MSM communities.

242 The implementation of public health campaigns and preventive measures, as well as

243 treatment interventions with the new direct-acting antivirals will allow the

244 development of strategies to reduce HCV transmission within these high-risk groups.

245

258

259   **Competing interests**

262

263 **References**

264 1 van de Laar TJW, van der Bij AK, Prins M, Bruisten SM, Brinkman K, Ruys TA, *et al.*
265   Increase in HCV incidence among men who have sex with men in Amsterdam
266   most likely caused by sexual transmission. *J Infect Dis* 2007; 196:230–238.

267 2 Laguno M, Martinez-Rebollar M, Perez I, Costa J, Larrousse M, Calvo M, *et al.*
268   Low rate of sustained virological response in an outbreak of acute hepatitis C in
269   HIV-infected patients. *AIDS Res Hum Retroviruses* 2012; 28:1294–1300.

270 3 Martínez-Rebollar M, Mallolas J, Pérez I, González-Cordón A, Loncà M, Torres B,
271   *et al.* Acute outbreak of hepatitis C in human immunodeficiency virus-infected
272   patients. *Enferm Infecc Microbiol Clin* 2015; 33:3–8.

273 4 Götz HM, van Doornum G, Niesters HG, den Hollander JG, Thio HB, de Zwart O,
274   *et al.* A cluster of acute hepatitis C virus infection among men who have sex with
275   men--results from contact tracing and public health implications. *AIDS* 2005;
276   19:969–974.

277 5 Vogel M, van de Laar T, Kupfer B, Stellbrink H-J, Kümmerle T, Mauss S, *et al.*
278   Phylogenetic analysis of acute hepatitis C virus genotype 4 infections among
279   human immunodeficiency virus-positive men who have sex with men in
280   Germany. *Liver Int* 2010; 30:1169–1172.

281 6 van de Laar T, Pybus O, Bruisten S, Brown D, Nelson M, Bhagani S, *et al.*
282   Evidence of a Large, International Network of HCV Transmission in HIV-Positive
283   Men Who Have Sex With Men. *Gastroenterology* 2009; 136:1609–1617.

284 7 Fierer DS, Fishman S, Uriel AJ, Carriero DC, Factor S, Dietrich DT, *et al.*
285   Characterization of an Outbreak of Acute HCV Infection in HIV-infected Men in
286   New York City. *16th Conf Retroviruses Opportunistic Infect* 2009; :poster 802.

287 8 Nishijima T, Shimbo T, Komatsu H. Incidence and Risk Factors for Incident
288   Hepatitis C Infection Among Men Who Have Sex With Men With HIV-1 Infection
289   in a Large Urban HIV Clinic in Tokyo. *J Acquir Immune Defic Syndr* 2014; 65:213–
290   217.

291 9 Matthews G, Hellard M, Haber P, Yeung B, Marks P, Baker D, *et al.*
292   Characteristics and treatment outcomes among HIV-infected individuals in the
293   Australian Trial in Acute Hepatitis C. *Clin Infect Dis* 2009; 48:650–658.

294 10 Pasquier C, Bujan L, Daudin M, Righi L, Berges L, Thauvin L, *et al.* Intermittent
295   detection of hepatitis C virus (HCV) in semen from men with human
296   immunodeficiency virus type 1 (HIV-1) and HCV. *J Med Virol* 2003; 69:344–349.

297 11 Urbanus AT, van de Laar TJ, Stolte IG, Schinkel J, Heijman T, Coutinho R a, *et al.*
298   Hepatitis C virus infections among HIV-infected men who have sex with men: an
299   expanding epidemic. *AIDS* 2009; 23:F1–F7.

300 12 Messina JP, Humphreys I, Flaxman A, Brown A, Cooke GS, Pybus OG, *et al.*
301   Global distribution and prevalence of hepatitis C virus genotypes. *Hepatology*
302   2015; 61:77–87.

303 13 De Bruijne J, Schinkel J, Prins M, Koekkoek SM, Aronson SJ, Van Ballegooijen

304   MW, *et al.* Emergence of hepatitis C virus genotype 4: Phylogenetic analysis
305   reveals three distinct epidemiological profiles. *J Clin Microbiol* 2009; 47:3832–
306   3838.

307 14 Quer J, Gregori J, Rodríguez-Frias F, Buti M, Madejon A, Perez-del-Pulgar S, *et al.*
308   High-Resolution Hepatitis C Virus Subtyping Using NS5B Deep Sequencing and
309   Phylogeny, an Alternative to Current Methods. *J Clin Microbiol* 2015; 53:219–
310   226.

311 15 Gregori J, Esteban JI, Cubero M, Garcia-Cehic D, Perales C, Casillas R, *et al.* Ultra-
312   deep pyrosequencing (UDPS) data treatment to study amplicon HCV minor
313   variants. *PLoS One* 2013; 8. doi:10.1371/journal.pone.0083361

314 16 van Asten L, Verhaest I, Lamzira S, Hernandez-Aguado I, Zangerle R, Boufassa F,
315   *et al.* Spread of hepatitis C virus among European injection drug users infected
316   with HIV: a phylogenetic analysis. *J Infect Dis* 2004; 189:292–302.

317 17 van de Laar TJW, Langendam MW, Bruisten SM, Welp E a E, Verhaest I, van
318   Ameijden EJC, *et al.* Changes in risk behavior and dynamics of hepatitis C virus
319   infections among young drug users in Amsterdam, the Netherlands. *J Med Virol*
320   2005; 77:509–18.

321 18 Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H,
322   *et al.* Clustal W and Clustal X version 2.0. *Bioinformatics* 2007; 23:2947–2948.

323 19 Guindon S, Dufayard J-F, Lefort V, Anisimova M, Hordijk W, Gascuel O. New
324   algorithms and methods to estimate maximum-likelihood phylogenies: assessing
325   the performance of PhyML 3.0. *Syst Biol* 2010; 59:307–321.

326 20 Posada D, Crandall KA. MODELTEST: testing the model of DNA substitution.
327   *Bioinformatics* 1998; 14:817–818.

328 21 Felsenstein J. Confidence limits on phylogenies: an approach using the bootstrap.
329   *Evolution (N Y)* 1985; 39:783–791.

330 22 Excoffier L, Lischer HEL. Arlequin suite ver 3.5: A new series of programs to
331   perform population genetics analyses under Linux and Windows. *Mol Ecol*
332   *Resour* 2010; 10:564–567.

333 23 Montoya-Ferrer A, Fierer DS, Alvarez-Alvarez B, de Gorgolas M, Fernandez-
334   Guerrero ML. Acute hepatitis C outbreak among HIV-infected men, Madrid,
335   Spain. *Emerg Infect Dis* 2011; 17:1560–2.

336 24 Sánchez C, Plaza Z, Vispo E, de Mendoza C, Barreiro P, Fernández-Montero J V.,
337   *et al.* Scaling up epidemics of acute hepatitis C and syphilis in HIV-infected men
338   who have sex with men in Spain. *Liver Int* 2013; 33:1357–1362.

339 25 Nevot M, Boesecke C, Parera M, Andrés C, Franco S, Revollo B, *et al.* Hepatitis C
340   virus NS3/4A quasispecies diversity in acute hepatitis C infection in HIV-1 co-
341   infected patients. *J Viral Hepat* 2014; 21. doi:10.1111/jvh.12254

342 26 Ingiliz P, Krznaric I, Stellbrink HJ, Knecht G, Lutz T, Noah C, *et al.* Multiple
343   hepatitis C virus (HCV) reinfections in HIV-positive men who have sex with men:
344   No influence of HCV genotype switch or interleukin-28B genotype on
345   spontaneous clearance. *HIV Med* 2014; 15:355–361.

15

346 27 Ingiliz P, Martin TC, Rodger A, Stellbrink H-J, Mauss S, Boesecke C, *et al.* HCV
347   reinfection incidence and spontaneous clearance rates in HIV-positive men who
348   have sex with men in Western Europe. *J Hepatol* 2017; 66:282–287.

349 28 Grabowski MK, Redd AD. Molecular tools for studying HIV transmission in sexual
350   networks. *Curr Opin HIV AIDS* 2014; 9:126–133.

351 29 Lamoury FMJ, Jacka B, Bartlett S, Bull RA, Wong A, Amin J, *et al.* The Influence of
352   Hepatitis C Virus Genetic Region on Phylogenetic Clustering Analysis. *PLoS One*
353   2015; 10:e0131437.

354 30 Wang GP, Sherrill-Mix SA, Chang K-M, Quince C, Bushman FD. Hepatitis C virus
355   transmission bottlenecks analyzed by deep sequencing. *J Virol* 2010; 84:6218–
356   6228.

357 31 Sato M, Maekawa S, Komatsu N, Tatsumi A, Miura M, Muraoka M, *et al.* Deep
358   sequencing and phylogenetic analysis of variants resistant to interferon-based
359   protease inhibitor therapy in chronic hepatitis induced by genotype 1b hepatitis
360   C virus. *J Virol* 2015; 89:6105–16.

361 32 Gonçalves Rossi LM, Escobar-Gutierrez A, Rahal P. Multiregion deep sequencing
362   of hepatitis C virus: An improved approach for genetic relatedness studies.
363   *Infect Genet Evol* 2016; 38:138–145.

364 33 Montoya V, Olmstead A, Tang P, Cook D, Janjua N, Grebely J, *et al.* Deep
365   sequencing increases hepatitis C virus phylogenetic cluster detection compared
366   to Sanger sequencing. *Infect Genet Evol* 2016; 43:329–337.

367 34 Ramos-Sanchez MC, Torio-Cabezon R, Mazon-Ramos MA, Martin-Gil FJ, Del
368   Alamo M. Hepatitis C virus genotype 4 in a North-west Spain district. *J Clin Virol*
369   2003; 28:223–224.

370 35 Bruguera M, Forns X. Hepatitis C en España. *Med Clin (Barc)* 2006; 127:113–117.

371 36 Pérez-Olmeda M, Ríos P, Núñez M, García-Samaniego J, Romero M, Soriano V.
372   Virological characteristics of hepatitis C virus infection in HIV-infected
373   individuals with chronic hepatitis C: implications for treatment. *AIDS* 2002;
374   16:493–5.

375

376

Figure Legends

378

379 **Fig. 1.** Maximum likelihood trees were constructed for each genotype separately: (a)

380 gt1a, (b) gt1b and (c) gt4d. Only sequences with relative frequency >0.5% in the

381 corresponding sample were included in the analyses. AHC HIV-coinfected patients are

382 identified as pXX, where XX represents a number from 1 to 84, and control patients are

383 identified as kXXXX, were XXXX represents a 4 digit number. Bootstrap support values

384 >80% are indicated in the corresponding nodes. Triangle areas are proportional to the

385 diversity and frequency of the different haplotypes (quasiespecies) in a sample. Capital

386 letters (black) indicate potential transmission clusters analyzed in more detail in

387 Supplementary Table S4.

388 **Fig. 2.** Maximum likelihood tree constructed with HCV gt4d sequences derived from

389 the international network reported by Van de Laar et al. (2009) and from our study

390 cohort of AHC HIV-coinfected MSM. Patients in our study cohort are identified as pXX,

391 for HAC (red), or kXXXX, for local controls (brown). Patients from the international

392 network were denoted as reported: FP, France (blue); ELR, ELC and EB, United

393 Kingdom (grey); S, NR, NA and PP, The Netherlands (purple); GB, Germany (green).

394 Nodes with bootstrap >80% are indicated in the tree.

Table 1

Table 1. Patient characteristics at the time of AHC diagnosis.

	AHC cases (n= 84)
Age (years)	39 (34-44)
Years between HIV and HCV diagnoses	5.6 (0.1-8.6)
Baseline CD4 (cells/μL)	564 (454-742)
Baseline % CD4	29 (21-34)
Baseline HIV viral load (IU/mL)	Undet*
Antiretroviral therapy	64 (76%)
Symptoms	23 (27%)
HCV-RNA (Log IU/mL)	6.28 (4.78-6.66)
AST (IU/L)	191 (102-329)
ALT (IU/L)	460 (224-702)
HCV Genotype:	
1a	33 (39%)
1b	7 (8%)
3a	1 (1%)
4d	43 (51%)
Other sexually transmitted diseases	36 (43%)
Interval diagnoses-start therapy (months)	2.0 (1.5-2.6)
PegIFN therapy	66 (79%)
Sustained virological response	45 (68%)

Quantitative variables are shown as median (range) and qualitative variables are shown as n (%).

Undet, Undetectable HIV viral load (Limit of detection= 37 IU/mL)

* One patient had very low HIV viral load of 67 IU/mL.

Fig. 1(a) HCV gt1a
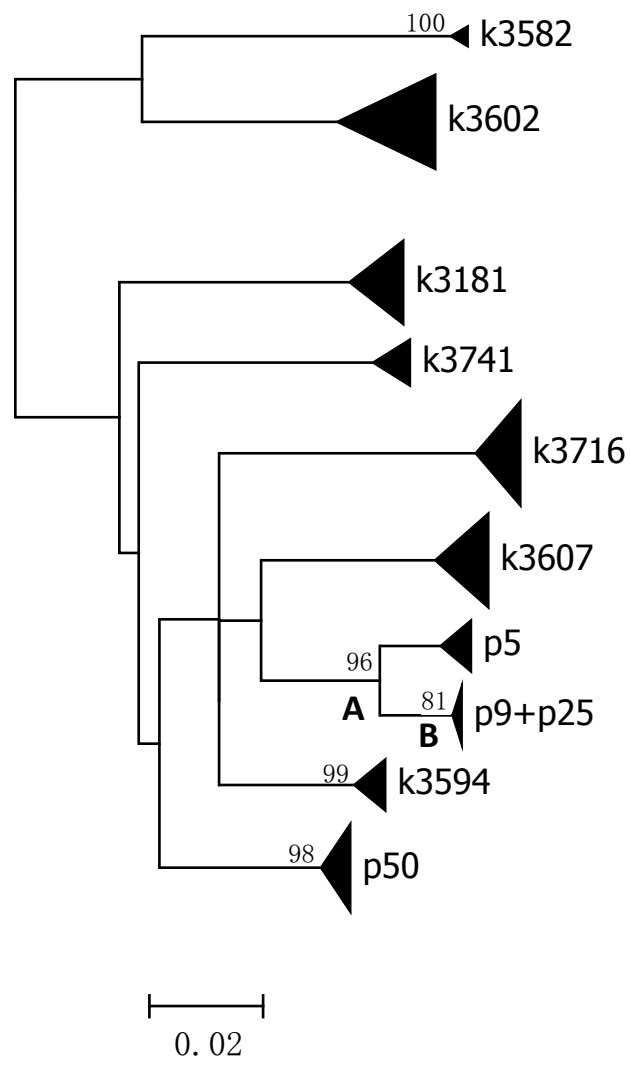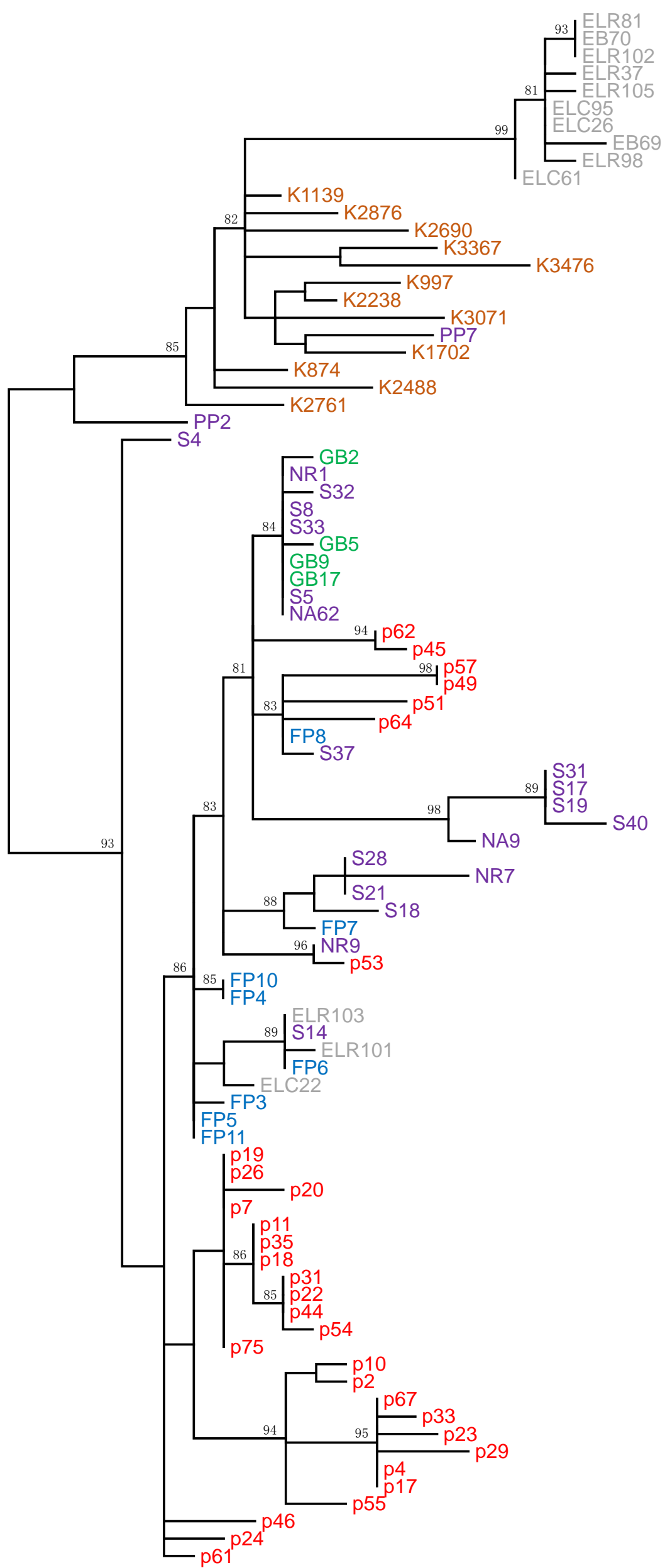
			97	p1
	86	**A**		
			92	p59
			k0417	

A
97 — p1
86
92 — p59
k0417

81 — **P**
88 — **B** — p14+p60
94 — **C** — p3+p63

87 — **O**
99 — p83
99 — p27

97 — **F**
89 — **D** — p30+p41
E — p37+p78+p84

96 — **G** — p28+p58
k3786

k4213
99 — p6

k3523
k3098

84
91 — **H** — p21+p32+p39+p40

k4572
91 — **I** — p56+p68+p79+p80

91 — **K** — p66+p72
93 — p73
95 — **L** — **J**
p70
M
84 — k2084
89 — p71

87 — k4157

97 — p43

97
98 — p65
N
98 — p69

0. 01

Fig. 1(b) HCV gt1b

Fig. 1(c) HCV gt4d

Fig. 2

HCV-NS5B amplification for ultra-deep pyrosequencing

Amplification and ultra-deep pyrosequencing of a 340 bp fragment of the HCV-NS5B region (nt 8279 to 8618, according to the isolate H77, data available in Genbank under accession number AF009606) was performed as previously described (1). HCV-RNA was extracted from 140 µL of serum by manual RNA extraction using the QIAamp viral RNA minikit (Qiagen, Hilden, Germany), as specified by the manufacturer. RT-PCR was performed using Transcriptor One Step RT-PCR Kit (Roche Applied Science Basel, Switzerland), using 20 pmol of sense primer 5Bo8254 and antisense primer 5Bo8707. Reverse transcription was performed at 50ºC for 30 min followed by a 2-step PCR reaction including: denaturing for 7 min at 94ºC; a first step of 10 cycles of 10 s at 94ºC, 30 s at 50ºC, and 1 min at 68ºC; a second step of 25 cycles of 10 s at 94ºC, 30 s at 50ºC and 1 min at 68ºC, increasing extension time 5 s per cycle; a final 7 min step at 68ºC.

Hemi-nested-PCR was performed using Fast Start High Fidelity PCR System, dNTPack (Roche Applied Science Basel, Switzerland). Briefly, 5µL from the previous PCR were amplified by a second PCR using sense primer 13n5Bo8254 and antisense primer 13n5Bo8641. These primers are composed by universal M13 forward and M13 reverse at 5' ends followed by a specific fragment complementary of the HCV PCR product amplified in the first round (see List of Primers below). Nested-PCR conditions were the following: denaturing for 2 min at 95ºC; 35 cycles of 30 s at 95ºC, 30 s at 55ºC and 1 min at 72ºC; a final 7 min step at 72ºC. For sample identification, the final product of the hemi-nested

amplification, was subjected to 15-cycles of re-amplification using primers composed by a complementary universal M13 primer (either sense or antisense) followed by a Roche's Validated Multiplex Identifier (MID) and with oligo A or B at 5' or 3' end of the sense or antisense primer, respectively.

Ultra-deep pyrosequencing and data treatment

Massive sequencing was performed in the GS-Junior 454/Roche platform (Roche, Branford, CT, USA), using titanium chemistry, which enables sequencing of 400- to 500-nt fragments (GS Junior Titanium Sequencing Kit), following manufacturer recommendations. The data used for the analysis was the FASTA files obtained from the 454 GS Junior system's software, which applies stringent quality controls on each sequenced nucleotide to guarantee the integrity of the full length of the amplicon. Briefly, the sequences were first demultiplexed by identifying MID and specific primer for each strand, and quality filtered by excluding all haplotypes with more than two Ns (any base), three gaps, not covering the full amplicon, or with an identity below 67% relative to the master sequence (defined here as the most abundant haplotype in the corresponding population). The accepted haplotypes with Ns and/or gaps were repaired by comparison with the dominant haplotype. Data treatment was performed using the pipelines designed to obtain clean nucleotide haplotypes (2) and for HCV subtyping (1).

HCV-NS5B amplification for Sanger direct sequencing

HCV-RNA was purified as described above. HCV-NS5B amplification of a 430 bp fragment of the HCV-NS5B region (nt 8553 to 8982 according to the isolate H77, data available in Genbank under accession number AF009606) was performed as described elsewhere (3,4) with some modifications. Briefly, cDNA was synthesized using genotype-specific primer 4HCV-OA (see List of Primers below) and AMV at 42ºC for 60 minutes in the presence of RNAsin (Pomega, Madison, WI). 5 µl of cDNA were used as template in a first-round PCR containing NS5B outer primers 4HCV-OS (sense) and 4HCV-OA (antisense), and 1.05 U of Expand High Fidelity Taq polymerase (Roche Applied Science). PCR cycling conditions were: 5 min at 95ºC, 35 cycles at 95ºC for 30 s, 50ºC for 30 s and 72º C for 1 min and a final extension of 7 min at 72ºC. 5 µl of the first-round product were then used in second-round reactions, using identical conditions as for the first-round and inner primers 4HCV-IS (sense) and E1b (I/A) (antisense).

PCR products were purified using the QIAquick PCR purification kit (Qiagen, Hilden, Germany). In order to obtain the population sequence (direct sequence), 20-50 ng of purified PCR product were sequenced bi-directionally using the set of primers of the corresponding nested (inner) PCR and the Big Dye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems, Foster City, CA, USA).

All extractions and amplifications were run according to universally adopted precautions, such as the use of different rooms for pre-PCR experiments and post-PCR experiments, in order to avoid cross-contamination. Negative controls for each step of RNA extraction and amplification were included.

List of Primers

Primer ID	Primer Sequence (5'→3')	Nucleotide Position*
5Bo8254	CNTAYGAYACCMGNTGYTTTGACTC	8254-8278
5Bo8707	TTNGADGAGCADGATGTWATBAGCTC	8682-8707
13n5Bo8254	*GTTGTAAAACGACGGCCAGT*CNTAYGAYACCMGNTGYTTTGACTC	8256-8278
13n5Bo8641	*CACAGGAAACAGCTATGACC*GARTAYCTGGTCATAGCNTCCGTGAA	8619-8641
4HCV-OS	ACCACCAGCTTYGGRAACAC	8457-8476
4HCV-OA	TTCGTGTGGAGAGTATCCRTGCA	9022-9044
4HCV-IS	CTGAGAGACTGCACSATGYTGGT	8523-8545
E1b (I/A)	AATGCGCTRAGRCCATGGAGTC	8995-9016

*Nucleotide position according to isolate H77, accession number AF009606
Primer sequences in italics correspond to M13 universal primers

<u>**References**</u>

1. Quer J, Gregori J, Rodríguez-Frias F, Buti M, Madejon A, Perez-del-Pulgar S, et al. High-Resolution Hepatitis C Virus Subtyping Using NS5B Deep Sequencing and Phylogeny, an Alternative to Current Methods. J Clin Microbiol. 2015;53(1):219–26.
2. Gregori J, Esteban JI, Cubero M, Garcia-Cehic D, Perales C, Casillas R, et al. Ultra-deep pyrosequencing (UDPS) data treatment to study amplicon HCV minor variants. PLoS One. 2013;8(12).
3. van Asten L, Verhaest I, Lamzira S, Hernandez-Aguado I, Zangerle R, Boufassa F, et al. Spread of hepatitis C virus among European injection drug users infected with HIV: a phylogenetic analysis. J Infect Dis. 2004;189(2):292–302.
4. van de Laar TJW, Langendam MW, Bruisten SM, Welp E a E, Verhaest I, van Ameijden EJC, et al. Changes in risk behavior and dynamics of hepatitis C virus infections among young drug users in Amsterdam, the Netherlands. J Med Virol. 2005;77(4):509–18.

Supplementary Tables

Table S1. HCV reinfection episodes.

Episode	Cases involved	Date of diagnosis	HCV genotype
Episode 1	p5	02/05/2011	1b
	p23	12/11/2012	4d
Episode 2	p69	10/06/2009	1a
	p7	08/08/2011	4d
Episode 3	p25	01/06/2012	1b
	p49	05/09/2013	4d
Episode 4	p57	16/10/2013	4d
	p58	26/04/2011	1a
Episode 5	p61	10/06/2009	4d
	p50	10/05/2013	1b

Table S2. Number of reads (forward + reverse) per sample, belonging to haplotypes with abundance above 0.5%. Control patients are shaded in blue.

	HCV genotype 1a		
Case	**Reads**	**Case**	**Reads**
p1	1622	p66	4177
p3	1469	p68	1912
p6	1261	p69	1265
p14	11558	p70	961
p21	1524	p71	763
p27	620	p72	3161
p28	379	p73	7112
p30	952	p78	1093
p32	612	p79	4919
p37	773	p80	3070
p39	311	p83	4559
p40	572	p84	7522
p41	4267	K3523	2042
p43	4131	K3786	4228
p56	6008	K4157	2421
p58	3409	K4213	5825
p59	2287	K4572	11059
p60	1252	K0398	1235
p63	2311	K0417	2155
p65	5638	K2084	312

Cont. Table S2.

	HCV genotype 1b		
Case	**Reads**	**Case**	**Reads**
p5	3284	K3602	731
p9	804	K3607	358
p25	1463	K3716	404
p50	6701	K3741	330
K3582	4280	K3181	1627
K3594	1357		

Cont. Table S2.

HCV genotype 4d			
Case	**Reads**	**Case**	**Reads**
p2	889	p55	449
p4	1859	p57	3752
p7	701	p61	376
p10	822	p62	468
p11	429	p64	1969
p17	4872	p67	1318
p18	3519	p75	2735
p19	840	p77	9390
p20	313	p81	2828
p22	854	p82	927
p23	779	K0075	387
p24	358	K0874	2708
p26	521	K0997	1450
p29	898	K1139	412
p31	526	K1702	3330
p33	446	K2238	1641
p35	3141	K2457	710
p44	627	K2488	590
p45	697	K2690	686
p46	3741	K2761	887
p49	1717	K2876	699
p51	6859	K3071	4267
p53	1025	K3363	1371
p54	567	K3476	3262

Table S3. Intra-patient genetic diversity in chronic hepatitis C (local controls) and AHC-HIV co-infected patients (AHC).

Genotype	Patients	Average	Maximum	Minimum
1a	Controls (n=8)	0.00260	0.00944	0.00015
	AHC (n=32)	0.00112	0.00457	0.00007
1b	Controls (n=7)	0.00632	0.01035	0.00154
	AHC (n=4)	0.00150	0.00214	0.00022
4d	Controls (n=14)	0.00552	0.01165	0.00101
	AHC (n=34)	0.00076	0.00212	0.00000

Genetic differentiation was calculated as the number of nt substitutions per site.

Table S4. Total and net genetic differentiation (substitutions/site) among cases in groups of AHC samples. Non-overlapping clusters are shaded in blue.

HCV genotype 1a		Total diversity	Net diversity
Groups	**Cases included**		
A	p1, p59	0.02208	0.02137
B	p14, p60	0.00306	0.00194
C	p3, p63	0.00332	0.00277
D (subcluster)*	p30, p41	0.00089	0.00001
E (subcluster)*	p37, p78, p84	0.00169	0.00046
F	p30, p41, p37, p78, p84	0.00929	0.00805
G	p28, p58	0.00072	0.00055
H	p21, p32, p39, p40	0.00111	0.00023
I	p56, p68, p79, p80	0.00877	0.00760
J	p70, p73	0.00285	0.00247
K	p66, p72	0.00262	0.00037
L**	p70, p73, p66, p72	0.01112	0.00985
M	p70, p73, p66, p72, p71	0.01119	0.00996
N	P65, p69	0.01716	0.01667
O**	p1, p59, p14, p60, p3, p63, p30, p41, p37, p78,p84, p83, p27, p28, p58	0.03936	0.03835
P**	p1, p59, p14, p60, p3, p63	0.02282	0.02188

* Group F is divided in 2 subclusters: groups D and E. Thus, they should be considered as a single group of transmission.
** Groups L, O and P were incorporated in the net diversity analysis despite the fact that they included one control patient because a priori we cannot exclude its relationship with the transmission network.

HCV genotype 1b		Total diversity	Net diversity
Groups	**Cases included**		
A	p9, p25, p5	0.01160	0.01006
B (subcluster)	p9, p25	0.00076	0.00008

Cont. Table S4.

HCV genotype 4d		Total diversity	Net diversity
Groups	Cases included		
A	All HCV-4d AHC patients (n=34)	0.01581	0.01468
B	p2, p4, p10, p17, p23, p29, p33, p45, p49, p51, p55, p57, p62, p64, p67, p82	0.01626	0.01537
C	p2, p4, p10, p17, p23, p29, p33, p55, p67	0.00738	0.00651
D (subcluster)	p4, p17, p23, p29, p33, p55, p67	0.00399	0.00317
E (subcluster)	p2, p10	0.00516	0.00441
F	p22, p31, p44, p54, p81	0.00299	0.00225
G	p45, p51, p62, p64, p82	0.01016	0.00840
H (subcluster)	p45, p62, p82	0.00226	0.00155
I	p35, p46	0.00023	0.00001
J	p18, p61	0.00303	0.00104
K	p24, p77	0.00178	0.00103
L	p19, p26	0.00173	0.00140