



Research paper

Multiregion deep sequencing of hepatitis C virus: An improved approach for genetic relatedness studies

Livia Maria Gonçalves Rossi^{a,b,*}, Alejandro Escobar-Gutierrez^b, Paula Rahal^a^a Department of Biology, Institute of Bioscience, Language and Exact Science, São Paulo State University, São José do Rio Preto, Sao Paulo, Brazil^b Instituto de Diagnóstico y Referencia Epidemiológicos, Mexico City, Mexico

ARTICLE INFO

Article history:

Received 22 November 2015

Received in revised form 23 December 2015

Accepted 24 December 2015

Available online 28 December 2015

Keywords:

Hepatitis C virus

Next generation sequencing

Outbreak

Multiregion

Genetic relatedness

ABSTRACT

Hepatitis C virus (HCV) is a major public health problem that affects more than 180 million people worldwide. Identification of HCV transmission networks is of critical importance for disease control. HCV related cases are often difficult to identify due to the characteristic long incubation period and lack of symptoms during the acute phase of the disease, making it challenging to link related cases to a common source of infection. Additionally, HCV transmission chains are difficult to trace back since viral variants from epidemiologically linked cases are genetically related but rarely identical. Genetic relatedness studies primarily rely on information obtained from the rapidly evolving HCV hypervariable region 1 (HVR1). However, in some instances, the rapid divergence of this region can lead to loss of genetic links between related isolates, which represents an important challenge for outbreak investigations and genetic relatedness studies. Sequencing of multiple and longer sub-genomic regions has been proposed as an alternative to overcome the limitations imposed by the rapid molecular evolution of the HCV HVR1. Additionally, conventional molecular approaches required to characterize the HCV intra-host genetic variation are laborious, time-consuming, and expensive while providing limited information about the composition of the viral population. Next generation sequencing (NGS) approaches enormously facilitate the characterization of the HCV intra-host population by detecting rare variants at much lower frequencies. Thus, NGS approaches using multiple sub-genomic regions should improve the characterization of the HCV intra-host population. Here, we explore the usefulness of multiregion sequencing using a NGS platform for genetic relatedness studies among HCV cases.

© 2015 Elsevier B.V. All rights reserved.

1. Introduction

Globally, hepatitis C virus (HCV) affects more than 180 million people (Mohd Hanafiah et al. 2013), in addition to 3–4 million new infections per year (Alter 2007; Lavanchy 2009). HCV infection is one of the leading causes of chronic liver disease associated with end-stage cirrhosis and hepatocellular carcinoma (Lauer and Walker 2001; McHutchison and Bacon 2005).

HCV is a small single-stranded, positive polarity, enveloped virus belonging to the *Hepacivirus* genus within the *Flaviviridae* family (Smith et al. 2014). The RNA viral genome (~9.6 kb in length) contains a single open reading frame encoding for a long polyprotein that upon maturation by enzymatic cleavage originates three structural proteins and seven non-structural proteins (Chevaliez and Pawlotsky 2006; Stanley et al. 2007). The HCV RNA replication process is highly error prone (Moradpour et al. 2007), and so far seven major HCV genotypes and several sub-types have been identified (Smith et al. 2014). Introduction of point mutations by the RNA polymerase is the primary element

contributing to the high genetic variability of HCV. The HCV mutation rate *in vivo* is $\sim 2.5 \times 10^{-5}$ per nucleotide per genome replication (Ribeiro et al. 2012); however, higher estimates have also been reported (Cuevas et al. 2009).

Rapid detection of HCV outbreaks and implementation of proper disease control measures are crucial to prevent virus spread and provide adequate health care. However, HCV transmission networks are difficult to identify. The intricate patterns of HCV molecular evolution and silent onset of disease complicate the recognition of transmission events (Goncalves Rossi and Rahal 2014). The hypervariable region 1 (HVR1) is generally used to characterize the HCV intra-host population (Campo et al. 2014; Forbi et al. 2014), and to detect HCV transmission by assessing the genetic relatedness of HVR1 variants among infected patients (Campo et al. 2015; Cruz-Rivera et al. 2013; Escobar-Gutierrez et al. 2013; Escobar-Gutierrez et al. 2012; Gismondini et al. 2013; Rossi et al. 2015). However, there are intrinsic limitations imposed by the use of a rapidly evolving sub-genomic region (Preciado et al. 2014; Rossi et al. 2015). Over time, genetic links can be lost due to rapid sequence divergence, impairing outbreak investigation studies (Cruz-Rivera et al. 2013). Thus, we propose that sequencing of additional sub-genomic regions might aid to restore links between

* Corresponding author.

E-mail address: liv.rossi@yahoo.com (L.M. Gonçalves Rossi).

isolates. Here, deep sequencing of the HCV HVR1 supplemented with sequences originated from the NS5A region was used to establish relatedness among HCV cases.

2. Materials and methods

2.1. Clinical samples

HCV chronic cases, aged 34–58 years, were enrolled in this study. These patients were a sub-set of a larger cohort (Cruz-Rivera et al. 2013; Escobar-Gutierrez et al. 2012; Fonseca-Coronado et al. 2012). All patients were anti-HCV treatment-naïve reporting injection drug users (IDU) activity. Ethical review and informed consent approval were granted by the Ethical Committee of the reference laboratory in Mexico. Informed consent was obtained from all subjects. Plasma samples from all subjects were obtained and stored at -70°C until use. Patients' characteristics are summarized in Table 1.

2.2. Amplicon deep sequencing

The intra-host viral genetic variation in each patient was evaluated by deep amplicon sequencing of the HCV HVR1 and NS5A regions using the 454 GS Junior system. Viral RNA was extracted using the QIAamp viral RNA kit (QIAGEN, Valencia, CA), according to the manufacturer's recommendations. Extracted RNA was used as a template for the RT-PCR using the One Step RT-PCR kit (QIAGEN), and specific first round primers recommended by the Centers for Disease Controls and Prevention (CDC) for each sub-genomic region (Ramachandran et al. 2011). Amplification of independent samples and regions was carried out under the following conditions: 50°C for 60 min, 95°C for 15 min, followed by 40 cycles at 95°C for 30 s, 55°C for 20 s, 72°C for 60 s. Subsequently, each sample was amplified independently with fusion primers including the 454 primer key (A and B for forward and reverse primers, respectively), a multiple identifier (MID) and the specific primers (Table 2). Hybrid MID tagging was performed in such way that each sample was amplified with a unique combination of forward and reverse primers, allowing for a limited number of fusion primers (Table 3). Nested PCR reaction was carried out using $2\ \mu\text{l}$ of the original first round PCR product as template. PCR products were subsequently resolved and purified by agarose gel electrophoresis on SizeSelect e-gels (Invitrogen, Carlsbad, CA). The quality and quantity of individual amplicons were assessed on a 2100 bioanalyzer (Agilent Technologies, Santa Clara, CA) using a DNA 1000 kit (Agilent Technologies). Purified amplicons were mixed at equimolar concentrations.

Table 1
Patients' characteristics.

Patient identifier	Gender	Age (yr)	HCV genotype	Viral titer (IU/ml)	Number of haplotypes	
					HVR1	NS5A
P1	Male	40	1a	3,680,000	744	125
P2	Male	44	1a	2,090,000	67	51
P3	Male	56	1a	2,790,000	366	172
P4	Female	45	1a	7,900,000	194	128
P5	Male	58	1a	4,400,000	299	81
P6	Male	43	1a	7,190,000	28	20
P7	Female	36	1a	5,740,000	13	166
P8	Male	51	1a	7,300,000	153	92
P9	Male	57	1a	8,430,000	95	289
P10	Female	49	1a	5,230,000	713	74
P11	Male	50	1a	4,080,000	79	164
P12	Female	42	1a	9,500,000	313	55
P13	Male	53	1b	6,204,000	135	123
P14	Female	56	1b	7,860,000	1145	106
P15	Male	34	1b	8,620,000	56	42
P16	Male	37	1b	5,370,000	24	131

Table 2
Primer sequences.

First Round	HVR1-F1	TGGCTTGGGATATGATGATGAACT
	HVR1-R1	GCAGTCTGTGATGTGCCA
	NS5A-F1	TCATAGAGGCCAACCTCTGTG
	NS5A-R1	TCGACCATGACCCGTCGCTGAG
454 sequencing	FWD-HVR-MID-1	CGTATCGCCTCCCTCGGCCATCAGACGAGTGCCG TGGATATGATGATGAACTGGT
	FWD-HVR-MID-2	CGTATCGCCTCCCTCGGCCATCAGACGCTCGAC AGGATATGATGATGAACTGGT
	FWD-HVR-MID-3	CGTATCGCCTCCCTCGGCCATCAGACGCGCACT CGGATATGATGATGAACTGGT
	FWD-HVR-MID-4	CGTATCGCCTCCCTCGGCCATCAGACGACTGTA GGGATATGATGATGAACTGGT
	RVS-HVR-MID-1	CTATGCGCCTTCCAGCCGCTCAGACGAGTG CGTTTGTATGTCAGCTGCCGTTGGTGT
	RVS-HVR-MID-2	CTATGCGCCTTCCAGCCGCTCAGACGCTCGAC ATTGATGTGCCAGCTGCCGTTGGTGT
	RVS-HVR-MID-3	CTATGCGCCTTCCAGCCGCTCAGACGCGCACT CTTGATGTGCCAGCTGCCGTTGGTGT
	RVS-HVR-MID-4	CTATGCGCCTTCCAGCCGCTCAGACGACTGTA GTTGATGTGCCAGCTGCCGTTGGTGT
	FWD-NS5a-MID-1	CGTATCGCCTCCCTCGGCCATCAGACGAGTGCCG TAGTGGTATTCTGGACTCTTTCG
	FWD-NS5a-MID-2	CGTATCGCCTCCCTCGGCCATCAGACGCTCGAC AAGTGGTATTCTGGACTCTTTCG
	FWD-NS5a-MID-3	CGTATCGCCTCCCTCGGCCATCAGACGCGCACT CAGTGGTATTCTGGACTCTTTCG
	FWD-NS5a-MID-4	CGTATCGCCTCCCTCGGCCATCAGACGACTGTA GAGTGGTATTCTGGACTCTTTCG
	RVS-NS5a-MID-1	CTATGCGCCTTCCAGCCGCTCAGACGAGTG CGTCATGGAGGAATAGGACTCAGCGTC
	RVS-NS5a-MID-2	CTATGCGCCTTCCAGCCGCTCAGACGCTCGAC ACATGGAGGAATAGGACTCAGCGTC
	RVS-NS5a-MID-3	CTATGCGCCTTCCAGCCGCTCAGACGCGCACT CTATGGAGGAATAGGACTCAGCGTC
	RVS-NS5a-MID-4	CTATGCGCCTTCCAGCCGCTCAGACGACTGTA GCATGGAGGAATAGGACTCAGCGTC

Amplicon deep sequencing was performed on a 454/Roche GS Junior instrument (Roche Applied Science, Indianapolis, IN) using the titanium chemistry (Roche Applied Science). Denoising of data sets was carried out using flow clustering as implemented in QIIME (Caporaso et al., 2010; Quince et al., 2011).

2.3. Phylogenetic and median joining network (MJN) analyses

Only long reads covering the entire length of the amplicon, including MID at both ends, were analyzed. Multiple alignment was performed using MAFFT v7 (Katoh and Standley 2013). Sequence reads were then analyzed using the neighbor joining method as implemented in MEGA 5 (Tamura et al. 2011). Intra-host genetic variation was also analyzed using MJN analysis (Bandelt et al. 1999), as implemented in Network v4.6 (Fluxus Technology) as previously described (Escobar-Gutierrez et al. 2013).

Table 3
Hybrid MID tags.

		Forward primer			
		MID1	MID2	MID3	MID4
Reverse primer	MID1	P1	P2	P3	P4
	MID2	P5	P6	P7	P8
	MID3	P9	P10	P11	P12
	MID4	P13	P14	P15	P16

3. Results

3.1. Phylogenetic analysis

3.1.1. HVR1 analysis

Initially, we selected a sub-set of seven patients from our cohort from whom we have information about risk factors. Patients 1–7 were epidemiologically linked, belonging to the same IDU network. Additionally, nine unrelated cases, including 4 genotype 1b infected patients, were included in the study. All seven patients were HVR1 PCR positive, and genotyping showed that all strains belonged to genotype 1A. Deep sequencing was carried out targeting the HVR1 region from all seven related patients and a sub-group of non-related cases. Clean and corrected read sequences from all isolates were used to generate the

corresponding phylogenetic (Fig. 1A). Related patients did not share any HCV variants. In fact, nucleotide distances from patients 1–5 were significantly large (Table 4). As a consequence, all variants were considerably distant from each other (minimal distance ranging from 15 to 23), exceeding the genetic distance threshold (Campo et al. 2015), effectively preventing us from assigning any genetic link. Therefore, genetic links purely based on sequence data from HVR1 could not be established. Interestingly, patients 1–5 exhibited very large intra-host genetic variability (maximum distance ranging from 30 to 60), suggesting the presence of more than one viral lineage in each patient (Table 4). On the other hand, patients 6 and 7 displayed a rather homogeneous viral population (average genetic distances 3.2 and 1.8, respectively), and were not genetically close to all other five patients in this network (minimal distance ranging from 35 to 48).

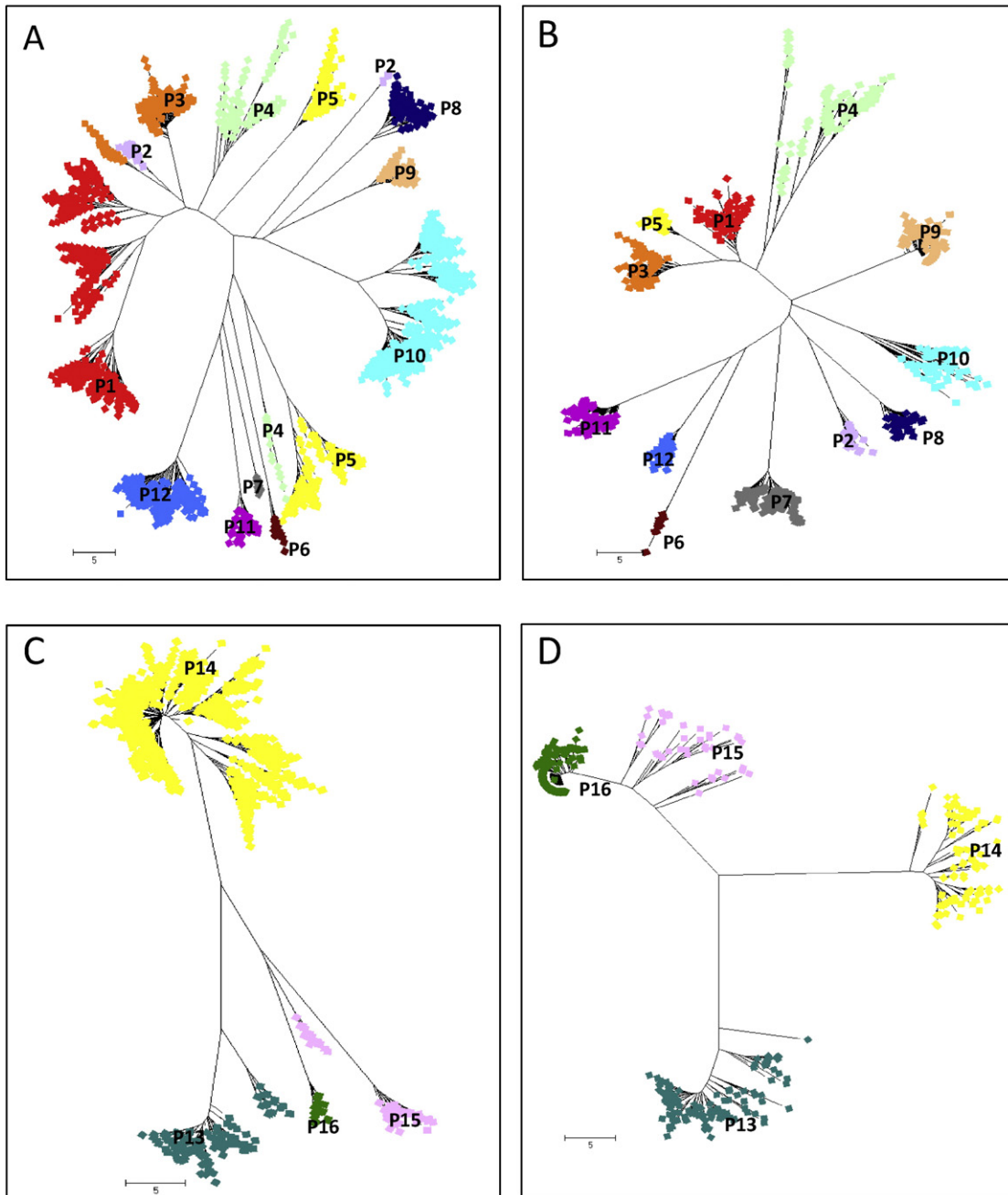


Fig. 1. Phylogenetic analysis. Phylogenetic trees were generated with clean and corrected sequence reads from the HVR1 (A and C) and NS5A (B and D) regions from all patients. The neighbor joining tree was constructed using maximum composite likelihood nucleotide distances using MEGA5 and 1000 replications for bootstrapping. Individual viral variants recovered from each individual patient are color coded.

Table 4
Intra- and inter-genetic distances for genotype 1a isolates.

		HVR1											NS5A																												
		P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12																
P1	Min	0.0											0.0																												
	Mean	18.6											4.5																												
	Max	35.0											14.0																												
P2	Min	14.0	0.0											32.0	1.0																										
	Mean	29.1	17.2											36.8	3.1																										
	Max	52.0	40.0											42.0	12.0																										
P3	Min	15.0	9.0	0.0										6.0	33.0	0.0																									
	Mean	27.7	25.4	8.5										12.4	37.2	3.2																									
	Max	45.0	53.0	30.0										22.0	44.0	10.0																									
P4	Min	18.0	12.0	16.0	0.0										6.0	31.0	13.0	1.0																							
	Mean	30.1	27.7	27.3	13.9										27.6	40.6	33.5	19.0																							
	Max	62.0	55.0	54.0	57.0										49.0	52.0	52.0	55.0																							
P5	Min	23.0	23.0	22.0	23.0	1.0										8.0	37.0	9.0	0.0																						
	Mean	43.1	44.0	40.9	42.9	28.7										12.2	41.0	12.8	32.3	2.2																					
	Max	63.0	62.0	57.0	63.0	60.0										18.0	45.0	18.0	51.0	5.0																					
P6	Min	47.0	48.0	45.0	42.0	42.0	0.0										48.0	44.0	49.0	43.0	52.0	0.0																			
	Mean	55.0	55.9	50.7	50.2	52.5	3.2										52.6	48.1	53.0	53.3	55.6	2.5																			
	Max	63.0	67.0	57.0	59.0	63.0	10.0										61.0	56.0	60.0	63.0	62.0	7.0																			
P7	Min	42.0	45.0	37.0	40.0	35.0	44.0	0.0										39.0	33.0	40.0	37.0	43.0	50.0	0.0																	
	Mean	50.3	51.8	42.4	44.8	47.2	46.6	1.8										45.6	38.6	45.6	46.0	47.3	54.7	4.0																	
	Max	57.0	62.0	49.0	50.0	60.0	50.0	4.0										54.0	46.0	53.0	55.0	54.0	62.0	11.0																	
P8	Min	35.0	36.0	36.0	31.0	40.0	53.0	43.0	1.0										32.0	24.0	32.0	34.0	36.0	52.0	37.0	0.0															
	Mean	44.1	44.0	44.5	40.5	50.4	60.4	47.7	5.5										36.7	28.9	36.9	44.5	39.7	55.6	42.5	2.5															
	Max	52.0	54.0	52.0	62.0	60.0	65.0	53.0	13.0											42.0	35.0	43.0	52.0	43.0	62.0	50.0	7.0														
P9	Min	31.0	32.0	29.0	31.0	39.0	48.0	37.0	32.0	0.0										24.0	29.0	29.0	30.0	30.0	43.0	38.0	34.0	0.0													
	Mean	40.2	38.2	38.7	39.7	50.6	52.7	40.9	38.5	3.4										29.7	34.7	34.7	42.3	34.8	48.1	43.6	38.5	2.6													
	Max	49.0	43.0	46.0	50.0	61.0	63.0	64.0	43.0	44.0	6.0										37.0	41.0	40.0	52.0	39.0	56.0	50.0	44.0	8.0												
P10	Min	31.0	33.0	32.0	29.0	37.0	51.0	41.0	35.0	28.0	1.0										24.0	26.0	25.0	30.0	28.0	43.0	36.0	26.0	28.0	1.0											
	Mean	45.8	41.4	40.8	40.0	50.1	56.7	47.5	45.1	36.5	13.3										33.2	33.1	33.3	45.7	36.2	48.8	43.4	32.8	34.4	13.3											
	Max	59.0	51.0	52.0	61.0	63.0	64.0	54.0	56.0	45.0	31.0										43.0	43.0	44.0	60.0	44.0	60.0	55.0	42.0	44.0	30.0											
P11	Min	43.0	46.0	43.0	42.0	36.0	39.0	35.0	47.0	38.0	48.0	0.0										37.0	38.0	42.0	30.0	45.0	37.0	38.0	40.0	38.0	34.0	0.0									
	Mean	51.8	52.1	53.9	49.5	46.8	43.3	39.7	52.8	43.8	54.9	3.3										43.8	41.9	47.9	43.9	49.9	41.8	43.0	44.6	43.8	43.3	2.8									
	Max	61.0	58.0	60.0	56.0	59.0	46.0	44.0	58.0	48.0	61.0	7.0										52.0	52.0	55.0	57.0	56.0	50.0	49.0	50.0	52.0	52.0	10.0									
P12	Min	40.0	46.0	42.0	39.0	36.0	51.0	35.0	49.0	40.0	38.0	34.0	1.0	34.0	37.0	39.0	28.0	40.0	38.0	37.0	36.0	37.0	34.0	26.0	0.0																
	Mean	51.4	52.1	49.9	47.3	48.8	56.4	40.1	56.0	46.3	48.1	39.1	5.0	40.3	41.9	44.3	43.3	44.2	42.9	43.0	40.2	42.6	41.8	31.0	3.2																
	Max	63.0	58.0	60.0	59.0	61.0	63.0	45.0	65.0	53.0	58.0	45.0	14.0	47.0	50.0	50.0	53.0	49.0	50.0	50.0	45.0	48.0	50.0	37.0	9.0																

3.1.2. NS5A analysis

HCV NS5A has been proposed to be a valuable sub-genomic region with enough sequence information to link related cases (Goncalves Rossi and Rahal 2014; Preciado et al. 2014; Ramachandran et al. 2011; Rossi et al. 2015). The lower mutation rate of the NS5A gene might allow for a better identification of transmission events particularly in those where rapid HVR1 evolution has taken place (Goncalves Rossi and Rahal 2014; Preciado et al. 2014). Thus, we then proceeded to amplify a larger PCR fragment (~500 bp) targeting the NS5A region and subsequently deep sequencing all amplicons. The NS5A sequences that originated from patients 1, 3, 4 and 5 formed an independent cluster; however, no identical variants were identified. The nucleotide distances between the NS5A sequences were considerably smaller than the HVR1 counterparts (minimal distances ranging from 6 to 13). Patient 2 was not part of this cluster and exhibited a rather homogeneous (average distance 3.1), NS5A intra-host population (Table 4), despite displaying two different populations (average distance 17.2) when HVR1 sequences were analyzed.

It is also noteworthy that patient 10 also exhibited a highly divergent HVR1 population (average distance 13.3) with two distinctive communities (Fig. 1). Interestingly, NS5A viral variants were equally heterogeneous (Table 4).

Four non-related, genotype 1b isolates were also included as controls to rule out genotypic bias (Fig. 1C). All isolates were non-related. Three isolates exhibited HVR1 highly divergent sequences (patients 13, 14 and 15). NS5A sequences derived from these patients were slightly less divergent (Table 5). Only one isolate displayed limited HVR1 sequence variability (patient 16). NS5A sequences showed limited genetic variability (average distance 2.9).

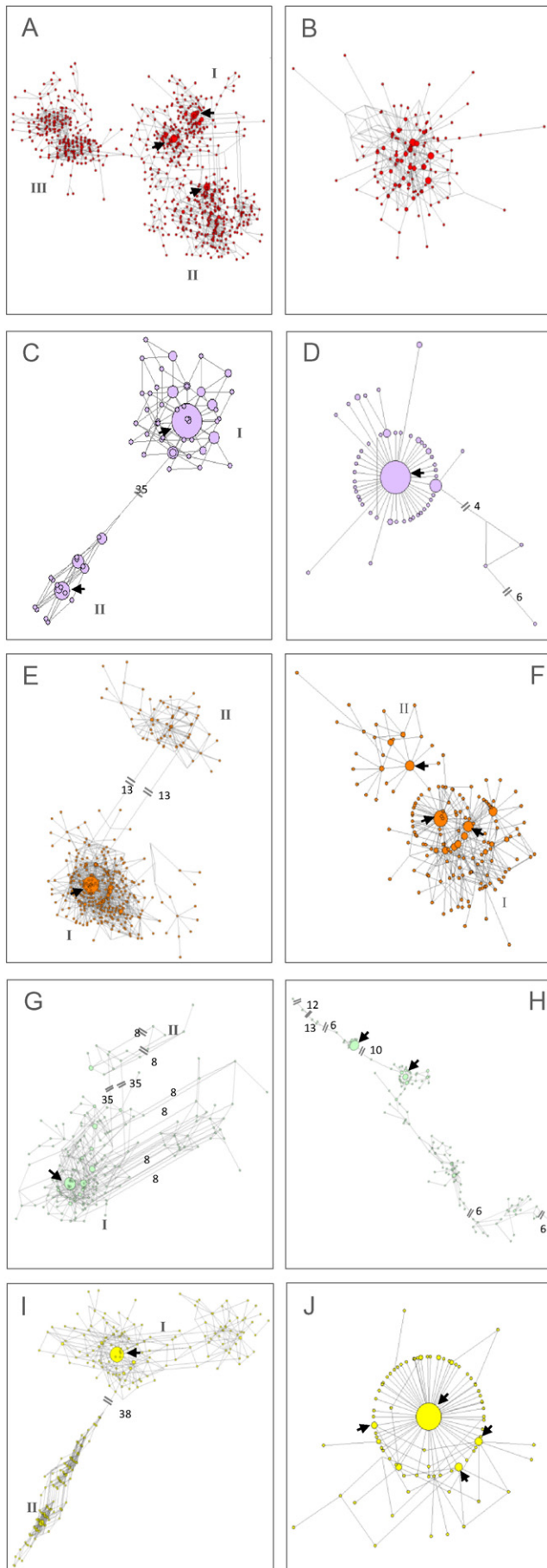
3.2. MJN analysis

3.2.1. Epidemiologically related, highly divergent isolates

MJN analysis was carried out to further analyze the intra-host genetic variability of all seven epidemiologically related patients (Fig. 2). As observed in the phylogenetic analysis, patient 1 exhibited three distinctive communities when HVR1 sequences were analyzed (Fig. 2A). The three viral communities were distant but genetically related. The main community (I) linked the other two communities and included most major viral variants, accounting for ~53.0% of all variants. The second community also included one of the main variants (8.0%). The third community was composed predominantly by low frequency

Table 5
Intra- and inter-genetic distances for genotype 1b isolates.

		HVR1				NS5A				
		P13	P14	P15	P16	P13	P14	P15	P16	
P13	Min	1.0				1.0				
	Mean	8.3				6.0				
	Max	25.0				21.0				
P14	Min	36.0	0.0			39.0	1.0			
	Mean	46.8	8.3			46.0	7.5			
	Max	57.0	24.0			55.0	22.0			
P15	Min	38.0	26.0		1.0		37.0	34.0	1.0	
	Mean	48.6	45.1	14.8		44.5	42.9	12.4		
	Max	55.0	57.0	34.0		53.0	52.0	26.0		
P16	Min	38.0	36.0	16.0	1.0		39.0	41.0	9.0	0.0
	Mean	43.5	44.4	30.0	2.6		46.1	46.4	15.6	2.9
	Max	48.0	51.0	38.0	5.0		52.0	52.0	23.0	7.0



variants and was the most distant population (Fig. 2A). Conversely, NS5A sequences derived from this isolate were significantly more homogeneous, grouping into a single community (Fig. 2B). Several major variants were identified within this community accounting for ~78.0% of all variants. HVR1 sequences from patient 2 displayed a different pattern (Fig. 2C). Two very distinctive communities, separated by 35 nucleotides, were identified. The main community contained most major variants that were closely related and separated mostly by 1-step mutations. The second community also included several major variants that were more diverse than the variants grouped in the main community. On the other hand, NS5A sequences derived from this patient were extremely homogeneous, including a major variant that accounted for 63.0% of all variants. Only three sequences were somehow distant (≥ 4 nucleotides), and far from the major variant (Fig. 2D). Patient 3 also presented two distinctive communities, 13 nucleotides apart, when HVR1 sequences were analyzed (Fig. 2E). The main community (I) included the major variant (47.0%) and the vast majority of all variants. The second community was scarcely populated and comprised primarily low frequency variants that were separated by 1-step mutations. NS5A viral variants derived from patient 3 seemed to group into two very closely related communities. The main community (I) included most of the major variants, although the second community also included few major variants. One of the major variants in community II was the linkage between both communities. Despite exhibiting two communities, patient 4 was mostly represented by variants contained in community I, including all major HVR1 viral variants (Fig. 2G). The minor community included only 14 HVR1 sequences that were somehow distant to such extent that could be separated into two further sub-communities. This was not surprising since NS5A sequences derived from the same patient indicated the presence of several non-related strains (Fig. 2H), possibly suggesting superinfection with multiple strains. Patient 5 also presented two fairly distinctive HVR1 communities (Fig. 2I). These communities were separated by at least 38 nucleotides. Community I included the major variant (48.0%) and an important number of low frequency closely related variants. Community II was more complex, basically composed by low frequency variants. Purely based on genetic distances, this second community could also be grouped into two different sub-communities. However, NS5A sequences derived from this patient were extremely homogenous forming a very compact network with one major viral variant at the center of the network surrounded by mostly minor variants, and few relative major variants, 1-step apart. Overall, NS5A sequences were less diverse than their HVR1 counterparts.

3.2.2. Epidemiologically related, non-divergent isolates

Patients 6 and 7 exhibited limited nucleotide diversity in both regions. The MJN analysis corroborated that observation and showed the existence of only one community in both isolates regardless of the sub-genomic region used to infer the sequence divergence (Fig. 3). Patient 7, however, exhibited a more complex NS5A population in comparison with HVR1 sequences (Table 4). While all variants derived from these patients still grouped into a single network, the complexity was significantly higher than the one derived from HVR1 sequences (Fig. 3D).

Fig. 2. MJN analysis for epidemiologically related and highly divergent isolates. MJN analyses from highly divergent isolates derived from epidemiologically related patients were performed as previously reported (Escobar-Gutierrez et al. 2013). The size of the node represents the percentage of that particular haplotype in the population. Long links were shortened and 'breaks' were inserted to facilitate visual display. The numbers above each break represent the number of nucleotides between the two nodes. Left and right panels depict HVR1 and NS5A MJN, respectively. Patient 1 (A and B), 2 (C and D), 3 (E and F), 4 (G and H) and 5 (I and J) are color coded as in Fig. 1. Arrows show major variants, and Roman numbers indicate communities within the viral population.

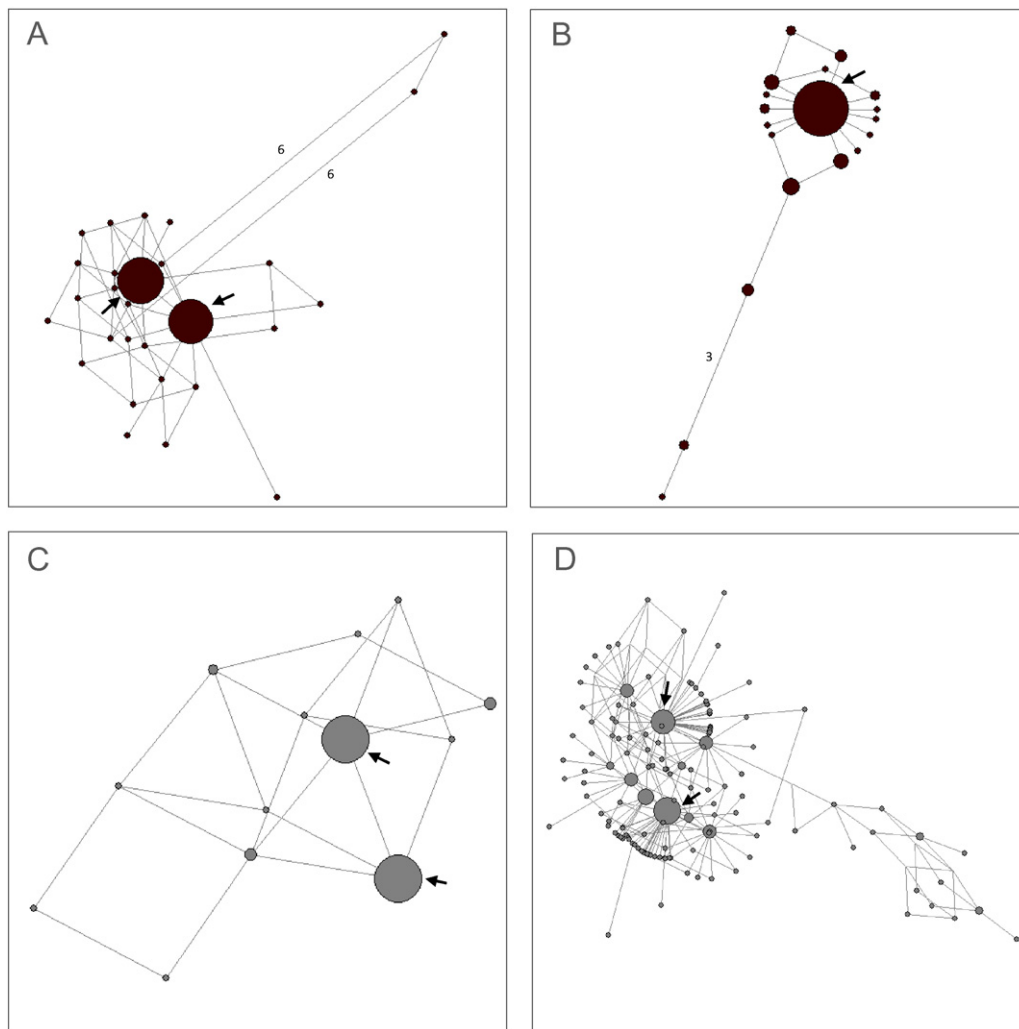


Fig. 3. MJN analysis for epidemiologically related and non-divergent isolates. MJN analyses from isolates with limited nucleotide diversity derived from epidemiologically related patients were performed as described previously (Escobar-Gutierrez et al. 2013). Left and right panels depict HVR1 and NS5A MJN, respectively. Patient 6 (A and B), 7 (C and D) are color coded as in Fig. 1. Arrows show major variants.

4. Discussion

Here, we have shown that multiregion sequencing significantly improves the resolution of HCV genetic relatedness studies. The results showed that deep sequencing paired with analysis of multiple regions is a powerful tool for molecular characterization of the viral population in clinical samples. Importantly, the use of both HVR1 and NS5A sequences in the phylogenetic analysis might help restore otherwise lost links when only information obtained from the HVR1 region alone is analyzed. Importantly, we were not able to link these cases using data from the NS5A region since accurate genetic thresholds are not currently established. However, our analysis is indicative of the circulation of closely related strains, likely over a long period of time, within the same IDU network.

Overall, NS5A sequences were less divergent than HVR1 sequences, even in those cases where multiple HVR1 viral communities were identified (Fig. 2B, D and J). This might indicate rapid intra-host evolution from an ancestor giving origin to lineages which still share closely related NS5A sequences. Alternatively, these might reflect superinfection with closely related viral lineages circulating in a close IDU network as reported by others (Cruz-Rivera et al. 2013). Our results cannot rule out either possibility. Thus, inter-host evolution could play a role in HCV intra-host genetic variability, since members of the same IDU network are likely to be exposed to related viral lineages

evolving over time within the network. Indeed, remarkably high intra-host genetic variability has been reported among IDU (Cruz-Rivera et al. 2013).

Interestingly, patients 3 and 4 exhibited more complex NS5A populations (Fig. 2F and H). This could be the reflection of introduction of relatively new viral lineages into the IDU network, leading to distinctive non-related communities within the same patient. In accordance with these observations, patients 6 and 7 displayed rather homogeneous viral populations, both HVR1 and NS5A (Fig. 3A and B). Despite not sharing sequences with patients 3 and 4 (Table 4), or being relatively close (minimal distance ≥ 37 for both sub-genomic regions), we cannot rule out the existence of other members of the same IDU network with related strains. Importantly, patients 6 and 7 also showed a limited genetic variability regardless of the region analyzed (Table 4). These might suggest a relative recent transmission event in these individuals. Genetic variability has been previously reported to be a dependable indicator of chronicity (Astrakhantseva et al. 2011; Montoya et al. 2015). Thus, in addition to aid establishing genetic relatedness, sequences derived from both regions can also provide information on incident cases related to recent transmission events.

The addition of a second informative region containing sufficient sequence information, such as the NS5A, in genetic relatedness studies might aid in restoring links among related cases. While the NS5A gene is relatively more conserved than the HVR1 (Rispetter et al.

2000), it still exhibits sufficiently genetic variability allowing its use in genetic relatedness studies. Our phylogenetic analysis derived from HVR1 sequences was incapable of linking cases since minimum genetic distances between cases significantly exceeded the genetic threshold (Campo et al. 2015). However, the inclusion of the NS5A region in the analysis brought the minimum sequence identity to a degree of closeness, suggesting that these HCV strains are circulating within the same IDU network.

Several limitations are imposed upon the use of rapidly evolving sub-genomic regions such as the HCV HVR1 for genetic relatedness studies (Cruz-Rivera et al. 2013; Rossi et al. 2015). Frequently, viral genomes sampled from cases and suspected sources are genetically related but rarely identical (Vandamme and Pybus 2013). The sequences originated from such cases are, however, likely more similar than those from non-epidemiologically related cases (Vandamme and Pybus 2013). In outbreak investigation, genetic closeness should be established by including suitable unrelated controls (Escobar-Gutierrez et al. 2012; Fischer et al. 2010); however, determining an appropriate genetic threshold is challenging (González-Candelas et al. 2013; Vandamme and Pybus 2013). The most important limitation of our study is the lack of a larger number of unrelated cases that would allow us to establish a reliable genetic threshold for the NS5A region. As reported previously, determining such genetic threshold is region and setting dependent (Campo et al. 2015; González-Candelas et al. 2013; Vandamme and Pybus 2013). Thus, larger studies are required to further refine our genetic analysis and adequately assigned relatedness links between HCV cases.

Intra-host genetic divergence in the HVR1 can reach up to 17%, while the inter-host divergence in unrelated cases can be as low as 5% (Bruno et al. 2015), further complicating the establishment of transmission links. In this study, the intra-host variability ranged between 0.0 and 60.0. The difficulty in establishing the genetic relationship between associated cases can be due to different evolutionary processes imposed on the HCV intra-host population during the course of infection (Rossi et al. 2015). Patients may experience a rapid evolution over short periods of time imposed by external factors such as antiviral therapy, or the result of potential re-infection and/or superinfection (Cruz-Rivera et al. 2013; Gonçalves Rossi and Rahal 2014; Rossi et al. 2015). Rapid sequence divergence significantly affects genetic relatedness between isolates associated with common sources of infection (Cruz-Rivera et al. 2013; Kurosaki et al. 1993; Ray et al. 2005), resulting in loss of links over time. In rapid evolving HCV populations, the high degree of divergence will likely cause an overlap of minimal genetic distances to those obtained from unrelated cases, thus preventing the correct assignment of cases to clusters of transmission (Cruz-Rivera et al. 2013). Moreover, phylogenetic analyses of the E1/E2 region not always faithfully reflect epidemiological relationships between isolates from serial transmissions (Casino et al., 1999). Additionally, changes in viral populations have been observed in patients undergoing combination peg IFN–RBV antiviral therapy, where patients showed different patterns of HVR1 evolution. In some cases, the majority of viral variants identified at baseline remained unchanged or showed little variation during therapy; however, in others, most baseline variants disappeared and were replaced by new variants (Saludes et al. 2013). In chronic cases, where evolution took place over a long period, the HCV population may be replaced by a new community (Gismondi et al. 2013). Nevertheless, constraints are sometimes observed in the HVR1 region, limiting its molecular evolution over time and maintaining the viral population relatively unaltered, even during longer periods of time (Gismondi et al. 2009; Larouche et al. 2012).

Ultra-deep sequencing is a powerful technology that allows for cost-effective probing of virus populations at an unprecedented level of detail. The massively parallel sequencing approach can detect low frequency mutations and provides a comprehensive picture of the entire virus population (Beerenwinkel and Zagordi 2011). Several studies have shown the advantages of using next generation sequencing (NGS) platforms over conventional methods for the analysis of the

HCV intra-host variability and molecular evolution (Bull et al. 2011; Fonseca-Coronado et al. 2012; Hiraga et al. 2011; Li et al. 2015; Merani et al. 2011; Montoya et al. 2015; Verbinen et al. 2010; Wang et al. 2010). Thus, amplicon deep sequencing is an appropriate technology for a comprehensive evaluation of the viral population in clinical samples (Campo et al. 2015; Escobar-Gutierrez et al. 2012; Lauck et al. 2012). Deep sequencing characteristic high-throughput yields millions of reads that allows the simultaneous sequencing of several individual samples, each tagged with a unique MID index. The costs associated with sample indexing could be further reduced by a simple solution. Here, each sample had a unique forward and reverse MID (Table 3), thus using only 4 sets of MID primers, we generated 16 MID tagging combinations. The only requirement for using this solution is that the sequence reads cover the entire amplicon length, since demultiplexing of individual samples is accomplished by searching for the corresponding combination of MIDs, instead of single indexes. In this study, each sub-genomic region was amplified and sequenced independently although simultaneous sequencing of both HVR1 and NS5A amplicons is possible (data not shown). Multiplex PCR is, however, more challenging since nested amplification is required. While theoretically possible, our experience multiplexing the primers initially reported by the CDC was not successful. Thus, further optimization is required to implement a multiplex PCR suitable for the simultaneous amplification of HVR1 and NS5A PCR fragments.

Using the 454 (Roche) platform, we have shown that longer and additional sub-genomic regions improved the establishment of genetic relationships among variants. Nonetheless, whole genome sequencing is the ultimate goal towards HCV molecular characterization and genetic relatedness studies. Although full-length viral haplotype reconstruction of heterogeneous population is computationally demanding, it has been shown to be feasible (Giallonardo et al. 2014; Bartolini et al., 2015). Besides haplotype reconstruction, further challenges will most probably be encountered. Low viral loads in many clinical specimens represent a major challenge for whole genome amplification; and therefore, more sensitive alternatives to conventional RT-PCR protocols are required. Moreover, the current NGS technologies have their own limitations, including shorter reads and in some cases lower accuracy. Undoubtedly, all platforms are able to generate usable sequences (Archer et al. 2012); however, their routine application on HCV outbreak investigation and whole genome sequencing remains to be explored.

In summary, HCV can exhibit different degrees of sequence divergence along its genome. These findings have important implications for outbreak investigations and genetic relatedness studies, as transmission links of related cases are lost over time due to different rates of viral evolution. These limitations can be overcome with the inclusion of other longer sub-genomic regions. Thus, the use of NGS technologies, and ultimately whole genome sequencing, will most likely further our understanding about the mechanisms exploited by HCV to spread in different settings as well as the molecular basis of HCV evolution.

Acknowledgments

LMGR was partially supported by a postgraduate fellowship from CAPES/DS, Brazil (n. 33004153079P9).

References

- Alter, M.J., 2007. Epidemiology of hepatitis C virus infection. *World J. Gastroenterol.* 13, 2436–2441.
- Archer, J., Weber, J., Henry, K., Winner, D., Gibson, R., Lee, L., Paxinos, E., Arts, E.J., Robertson, D.L., Mimms, L., Quinones-Mateu, M.E., 2012. Use of four next-generation sequencing platforms to determine HIV-1 coreceptor tropism. *PLoS One* 7, e49602.
- Astrakhantseva, I.V., Campo, D.S., Araujo, A., Teo, C.G., Khudyakov, Y., Kamili, S., 2011. Differences in variability of hypervariable region 1 of hepatitis C virus (HCV) between acute and chronic stages of HCV infection. *Silico Biol.* 11, 163–173.

- Bandelt, H.J., Forster, P., Rohlf, A., 1999. Median-joining networks for inferring intraspecific phylogenies. *Mol. Biol. Evol.* 16, 37–48.
- Bartolini, B., Giombini, E., Abbate, I., Selli, M., Rozera, G., Biagini, T., Visco-Comandini, U., Taibi, C., Capobianchi, M.R., 2015. Near full length hepatitis C virus genome reconstruction by next generation sequencing based on genotype-independent amplification. *Dig Liver Dis* 47 (7), 608–612.
- Beerenwinkel, N., Zagordi, O., 2011. Ultra-deep sequencing for the analysis of viral populations. *Curr. Opin Virol.* 1, 413–418.
- Bruno, S., Thompson, A.J., Critelli, R., Crosignani, A., Rossi, S., De Lisi, S., Cariani, E., Zermiani, P., Vaira, V., Boccaccio, V., Maisonneuve, P., Villa, E., 2015. Interferon lambda-3 is not associated with clinical outcome in patients with HCV-induced compensated cirrhosis: a long-term cohort study. *Antivir. Res.* 113, 27–32.
- Bull, R.A., Luciani, F., McElroy, K., Gaudieri, S., Pham, S.T., Chopra, A., Cameron, B., Maher, L., Dore, G.J., White, P.A., Lloyd, A.R., 2011. Sequential bottlenecks drive viral evolution in early acute hepatitis C virus infection. *PLoS Pathog.* 7, e1002243.
- Campo, D.S., Skums, P., Dimitrova, Z., Vaughan, G., Forbi, J.C., Teo, C.G., Khudyakov, Y., Lau, D.T., 2014. Drug resistance of a viral population and its individual intrahost variants during the first 48 hours of therapy. *Clin. Pharmacol. Ther.* 95, 627–635.
- Campo, D.S., Xia, G.L., Dimitrova, Z., Lin, Y., Forbi, J.C., Ganova-Raeva, L., Punkova, L., Ramachandran, S., Thai, H., Skums, P., Sims, S., Rytsareva, I., Vaughan, G., Roh, H.J., Purdy, M.A., Sue, A., Khudyakov, Y., 2015. Accurate Genetic Detection of Hepatitis C Virus Transmissions in Outbreak Settings. *J. Infect. Dis.*
- Caporaso, J.G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F.D., Costello, E.K., Fierer, N., Pena, A.G., Goodrich, J.K., Gordon, J.I., Huttley, G.A., Kelley, S.T., Knights, D., Koenig, J.E., Ley, R.E., Lozupone, C.A., McDonald, D., Muegge, B.D., Pirrung, M., Reeder, J., Sevinsky, J.R., Turnbaugh, P.J., Walters, W.A., Widmann, J., Yatsunenko, T., Zaneveld, J., Knight, R., 2010. QIIME allows analysis of high-throughput community sequencing data. *Nat. methods* 7, 335–336.
- Casino, C., McAllister, J., Davidson, F., Power, J., Lawlor, E., Yap, P.L., Simmonds, P., Smith, D.B., 1999. Variation of hepatitis C virus following serial transmission: multiple mechanisms of diversification of the hypervariable region and evidence for convergent genome evolution. *J. Gen. Virol* 80 (Pt 3), 717–725.
- Chevaliez, S., Pawlotsky, J.M., 2006. hcv genome and life cycle. In: Tan, S.L. (Ed.), *Hepatitis C Viruses: Genomes and Molecular Biology*. Norfolk (UK).
- Cruz-Rivera, M., Carpio-Pedroza, J.C., Escobar-Gutierrez, A., Lozano, D., Vergara-Castaneda, A., Rivera-Osorio, P., Martinez-Guarneros, A., Chacon, C.A., Fonseca-Coronado, S., Vaughan, G., 2013. Rapid hepatitis C virus divergence among chronically infected individuals. *J. Clin. Microbiol.* 51, 629–632.
- Cuevas, J.M., Gonzalez-Candelas, F., Moya, A., Sanjuan, R., 2009. Effect of ribavirin on the mutation rate and spectrum of hepatitis C virus in vivo. *J. Virol.* 83, 5760–5764.
- Escobar-Gutierrez, A., Vazquez-Pichardo, M., Cruz-Rivera, M., Rivera-Osorio, P., Carpio-Pedroza, J.C., Ruiz-Pacheco, J.A., Ruiz-Tovar, K., Vaughan, G., 2012. Identification of hepatitis C virus transmission using a next-generation sequencing approach. *J. Clin. Microbiol.* 50, 1461–1463.
- Escobar-Gutierrez, A., Soudeyns, H., Larouche, A., Carpio-Pedroza, J.C., Martinez-Guarneros, A., A., V.-C.C., Fonseca-Coronado, S., Yamasaki, L.H., Ruiz-Tovar, K., Cruz-Rivera, M.Y., 2013. Vertical Transmission of Hepatitis C Virus: A Tale of Multiple Outcomes. *Infect Genet Evol.*
- Fischer, G.E., Schaefer, M.K., Labus, B.J., Sands, L., Rowley, P., Azzam, I.A., Armour, P., Khudyakov, Y.E., Lin, Y., Xia, G., Patel, P.R., Perz, J.F., Holmberg, S.D., 2010. Hepatitis C virus infections from unsafe injection practices at an endoscopy clinic in Las Vegas, Nevada, 2007–2008. *Clin. Infect. Dis.* 51, 267–273.
- Fonseca-Coronado, S., Escobar-Gutierrez, A., Ruiz-Tovar, K., Cruz-Rivera, M.Y., Rivera-Osorio, P., Vazquez-Pichardo, M., Carpio-Pedroza, J.C., Ruiz-Pacheco, J.A., Cazares, F., Vaughan, G., 2012. Specific detection of naturally occurring hepatitis C virus mutants with resistance to telaprevir and boceprevir (protease inhibitors) among treatment-naïve infected individuals. *J. Clin. Microbiol.* 50, 281–287.
- Forbi, J.C., Campo, D.S., Purdy, M.A., Dimitrova, Z.E., Skums, P., Xia, G.L., Punkova, L.T., Ganova-Raeva, L.M., Vaughan, G., Ben-Ayed, Y., Switzer, W.M., Khudyakov, Y.E., 2014. Intra-host diversity and evolution of hepatitis C virus endemic to Cote d'Ivoire. *J. Med. Virol.* 86, 765–771.
- Giallonardo, F.D., Topfer, A., Rey, M., Prabhakaran, S., Dupont, Y., Leemann, C., Schmutz, S., Campbell, N.K., Joos, B., Lecca, M.R., Patrignani, A., Daumer, M., Beisel, C., Rusert, P., Trkola, A., Gunthard, H.F., Roth, V., Beerenwinkel, N., Metzner, K.J., 2014. Full-length haplotype reconstruction to infer the structure of heterogeneous virus populations. *Nucleic Acids Res.* 42, e115.
- Gismondi, M.I., Becker, P.D., Diaz Carrasco, J.M., Guzman, C.A., Campos, R.H., Preciado, M.V., 2009. Evolution of hepatitis C virus hypervariable region 1 in immunocompetent children born to HCV-infected mothers. *J. Viral Hepat.* 16, 332–339.
- Gismondi, M.I., Diaz Carrasco, J.M., Valva, P., Becker, P.D., Guzman, C.A., Campos, R.H., Preciado, M.V., 2013. Dynamic changes in viral population structure and compartmentalization during chronic hepatitis C virus infection in children. *Virology* 447, 187–196.
- Goncalves Rossi, L.M., Rahal, P., 2014. Challenges in molecular epidemiology of hepatitis C virus. *J. Clin. Virol.* 60, 174–176.
- González-Candelas, F., Bracho, M.A., Wróbel, B., Moya, A., 2013 Jul 19. Molecular evolution in court: analysis of a large hepatitis C virus outbreak from an evolving source. *BMC Biol.* 11, 76.
- Hiraga, N., Imamura, M., Abe, H., Hayes, C.N., Kono, T., Onishi, M., Tsuge, M., Takahashi, S., Ochi, H., Iwao, E., Kamiya, N., Yamada, I., Tateno, C., Yoshizato, K., Matsui, H., Kanai, A., Inaba, T., Tanaka, S., Chayama, K., 2011. Rapid emergence of telaprevir resistant hepatitis C virus strain from wildtype clone in vivo. *Hepatology* 54, 781–788.
- Katoh, K., Standley, D.M., 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* 30, 772–780.
- Kurosaki, M., Enomoto, N., Marumo, F., Sato, C., 1993. Rapid sequence variation of the hypervariable region of hepatitis C virus during the course of chronic infection. *Hepatology* 18, 1293–1299.
- Larouche, A., Gaetan, G., El-Bilali, N., Quesnel-Vallieres, M., Martin, S.R., Alvarez, F., Shoukry, N.H., Soudeyns, H., 2012. Seronegative hepatitis C virus infection in a child infected via mother-to-child transmission. *J. Clin. Microbiol.* 50, 2515–2519.
- Lauck, M., Alvarado-Mora, M.V., Becker, E.A., Bhattacharya, D., Striker, R., Hughes, A.L., Carrilho, F.J., O'Connor, D.H., Pinho, J.R., 2012. Analysis of hepatitis C virus intrahost diversity across the coding region by ultradeep pyrosequencing. *J. Virol.* 86, 3952–3960.
- Lauer, G.M., Walker, B.D., 2001. Hepatitis C virus infection. *N. Engl. J. Med.* 345, 41–52.
- Lavanthy, D., 2009. The global burden of hepatitis C. *Liver Int.* 29 (Suppl. 1), 74–81.
- Li, H., Stoddard, M.B., Wang, S., Giorgi, E.E., Blair, L.M., Learn, G.H., Hahn, B.H., Alter, H.J., Busch, M.P., Fierer, D.S., Ribeiro, R.M., Perelson, A.S., Bhattacharya, T., Shaw, G.M., 2015. Single-genome sequencing of hepatitis C virus in donor–recipient pairs distinguishes modes and models of virus transmission and early diversification. *J. Virol.* 90, 152–166.
- McHutchison, J.G., Bacon, B.R., 2005. Chronic hepatitis C: an age wave of disease burden. *Am. J. Manag. Care* 11, S286–S295 (quiz S307–211).
- Merani, S., Petrovic, D., James, I., Chopra, A., Cooper, D., Freitas, E., Rauch, A., di Iulio, J., John, M., Lucas, M., Fitzmaurice, K., McKiernan, S., Norris, S., Kelleher, D., Klenerman, P., Gaudieri, S., 2011. Effect of immune pressure on hepatitis C virus evolution: insights from a single-source outbreak. *Hepatology* 53, 396–405.
- Mohd Hanafiah, K., Groeger, J., Flaxman, A.D., Wiersma, S.T., 2013. Global epidemiology of hepatitis C virus infection: new estimates of age-specific antibody to HCV seroprevalence. *Hepatology* 57, 1333–1342.
- Montoya, V., Olmstead, A.D., Janjua, N.Z., Tang, P., Grebely, J., Cook, D., Richard Harrigan, P., Krajden, M., 2015. Differentiation of acute from chronic hepatitis C virus infection by nonstructural 5B deep sequencing: a population-level tool for incidence estimation. *Hepatology* 61, 1842–1850.
- Moradpour, D., Penin, F., Rice, C.M., 2007. Replication of hepatitis C virus. *Nat. Rev. Microbiol.* 5, 453–463.
- Preciado, M.V., Valva, P., Escobar-Gutierrez, A., Rahal, P., Ruiz-Tovar, K., Yamasaki, L., Vazquez-Chacon, C., Martinez-Guarneros, A., Carpio-Pedroza, J.C., Fonseca-Coronado, S., Cruz-Rivera, M., 2014. Hepatitis C virus molecular evolution: transmission, disease progression and antiviral therapy. *World J. Gastroenterol.* 20, 15992–16013.
- Quince, C., Lanzen, A., Davenport, R.J., Turnbaugh, P.J., 2011. Removing noise from pyrosequenced amplicons. *BMC bioinformatics* 12, 38.
- Ramachandran, S., Campo, D.S., Dimitrova, Z.E., Xia, G.L., Purdy, M.A., Khudyakov, Y.E., 2011. Temporal variations in the hepatitis C virus intrahost population during chronic infection. *J. Virol.* 85, 6369–6380.
- Ray, S.C., Fanning, L., Wang, X.H., Netski, D.M., Kenny-Walsh, E., Thomas, D.L., 2005. Divergent and convergent evolution after a common-source outbreak of hepatitis C virus. *J. Exp. Med.* 201, 1753–1759.
- Ribeiro, R.M., Li, H., Wang, S., Stoddard, M.B., Learn, G.H., Korber, B.T., Bhattacharya, T., Guedj, J., Parrish, E.H., Hahn, B.H., Shaw, G.M., Perelson, A.S., 2012. Quantifying the diversification of hepatitis C virus (HCV) during primary infection: estimates of the in vivo mutation rate. *PLoS Pathog.* 8, e1002881.
- Rispeter, K., Lu, M., Behrens, S.E., Fumiko, C., Yoshida, T., Roggendorf, M., 2000. Hepatitis C virus variability: sequence analysis of an isolate after 10 years of chronic infection. *Virus Genes* 21, 179–188.
- Rossi, L.M., Escobar-Gutierrez, A., Rahal, P., 2015. Advanced molecular surveillance of hepatitis C virus. *Viruses* 7, 1153–1188.
- Saludes, V., Gonzalez-Candelas, F., Planas, R., Sola, R., Ausina, V., Martro, E., 2013. Evolutionary dynamics of the E1–E2 viral populations during combination therapy in non-responder patients chronically infected with hepatitis C virus subtype 1b. *Infect. Genet. Evol.* 13, 1–10.
- Smith, D.B., Bukh, J., Kuiken, C., Muerhoff, A.S., Rice, C.M., Stapleton, J.T., Simmonds, P., 2014. Expanded classification of hepatitis C virus into 7 genotypes and 67 subtypes: updated criteria and genotype assignment web resource. *Hepatology* 59, 318–327.
- Stanley, M.L., Walker, C., Alter, M.J., Yi, M., 2007. Hepatitis C virus. In: Knipe, D.M., Howley, P.M. (Eds.), *Fields Virology*. Lippincott Williams & Wilkins, Philadelphia, p. 1280.
- Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., Kumar, S., 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol.* 28, 2731–2739.
- Vandamme, A.M., Pybus, O.G., 2013. Viral phylogeny in court: the unusual case of the Valencian anesthetist. *BMC Biol.* 11, 83.
- Verbinen, T., Van Marck, H., Vandenbroucke, I., Vijgen, L., Claes, M., Lin, T.I., Simmen, K., Neyts, J., Fanning, G., Lenz, O., 2010. Tracking the evolution of multiple in vitro hepatitis C virus replicon variants under protease inhibitor selection pressure by 454 deep sequencing. *J. Virol.* 84, 11124–11133.
- Wang, G.P., Sherrill-Mix, S.A., Chang, K.M., Quince, C., Bushman, F.D., 2010. Hepatitis C virus transmission bottlenecks analyzed by deep sequencing. *J. Virol.* 84, 6218–6228.