OPEN ACCESS Freely available online

PLoS one

Recombination in Hepatitis C Virus: Identification of Four Novel Naturally Occurring Inter-Subtype Recombinants

Weifeng Shi[1*¤], Ines T. Freitas[1◔], Chaodong Zhu[2], Wei Zheng[2,3,4], William W. Hall[5], Desmond G. Higgins[1]

1 The Conway Institute of Biomolecular and Biomedical Research, University College Dublin, Dublin, Ireland, **2** Key Laboratory of Zoological Systematics and Evolution, Institute of Zoology, Chinese Academy of Sciences, Beijing, China, **3** Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences, Shenzhen, China, **4** Graduate University of Chinese Academy of Sciences, Beijing, China, **5** National Virus Reference Laboratory, University College Dublin, Dublin, Ireland

Abstract

Recombination in Hepatitis C virus (HCV) is considered to be rare. In this study, we performed a phylogenetic analysis of 1278 full-length HCV genome sequences to identify potential recombination events. Nine inter-genotype recombinants were identified, all of which have been previously reported. This confirms the rarity of inter-genotype HCV recombinants. The analysis also identified five inter-subtype recombinants, four of which are documented for the first time (EU246930, EU246931, EU246932, and EU246937). Specifically, the latter represent four different novel recombination types (6a/6o, 6e/6o, 6e/6h, and 6n/6o), and this was well supported by seven independent methods embedded in RDP. The breakpoints of the four novel HCV recombinants are located within the NS5B coding region and were different from all previously reported breakpoints. While the locations of the breakpoints identified by RDP were not identical, they are very close. Our study suggests that while recombination in HCV is rare, this warrants further investigation.

Citation: Shi W, Freitas IT, Zhu C, Zheng W, Hall WW, et al. (2012) Recombination in Hepatitis C Virus: Identification of Four Novel Naturally Occurring Inter-Subtype Recombinants. PLoS ONE 7(7): e41997. doi:10.1371/journal.pone.0041997

Editor: John E. Tavis, Saint Louis University, United States of America

Received March 20, 2012; **Accepted** June 28, 2012; **Published** July 24, 2012

Funding: This study was supported by Science Foundation Ireland (PI grant 07/IN.1/B1783). I.T. Freitas was funded by the UK Wellcome Trust (097427/Z/11/Z). C. Zhu and W. Zheng were supported by the National Science Foundation, China (Grants Nos. 31172048, J0930004). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: wfshi.tsmc@gmail.com

¤ Current address: Guangzhou Institute of Advanced Technology, Chinese Academy of Sciences, Nansha, Guangzhou, China

◔ These authors contributed equally to this work.

Introduction

Hepatitis C Virus (HCV) belongs to the family *Flaviviridae* and was first identified in 1989 [1]. It is a major cause of the liver diseases: chronic hepatitis, cirrhosis, and hepatocellular carcinoma. HCV is an enveloped virus with a positive-sense, single-stranded RNA genome of approximately 9400 bp in length [2]. The HCV genome has one open reading frame encoding a polyprotein of about 3,000 amino acids, and this is processed to produce three structural (core, E1, E2) and seven non-structural proteins (p7, NS2, NS3, NS4A, NS4B, NS5A, NS5B) [3].

Similar to many RNA viruses, HCV exhibits high genetic heterogeneity and to date seven genotypes have been identified. Different genotypes diverge by at least 30% over the complete genome [4]. In addition, HCV has also been further classified into numerous subtypes (http://hcv.lanl.gov/content/sequence/HCV/classification/genotable.html). Subtypes can diverge by as much as 20%, but within subtype variation is usually less than 10% [4]. To date, genotype 1 includes 13 subtypes (subtypes 1a to 1m). The numbers of subtypes for genotypes 2, 3, and 4 were 18, 11, and 18, respectively. Genotypes 5 and 7 have only a single subtype, 5a and 7a. However, it is likely that more subtypes might be found for these genotypes due to continuous efforts to sequence more viral genomes. Genotype 6 has the largest number of reported subtypes with a total of 21.

Recombination is an important evolutionary process for many viruses, such as human immunodeficiency virus [5] and hepatitis B virus (HBV) [6,7]. However, recombination is considered to be rare in HCV [8,9]. This is supported by the finding that HCV-infected cells can rarely be superinfected by another HCV of a different group or subtype, *in vivo* [10]. However, HCV superinfection or co-infection is known to occur [11–15] and recombination, while rare, would be expected to occur.

Recently, Gonzalez-Candelas et al. classified HCV recombination events into three types: inter-genotype recombination, inter-subtype recombination, and intra-patient/intra-subtype recombination [9]. So far, seven inter-genotype recombination types (2k/1b, 2i/6p, 2b/1b, 2/5, 2b/6w, 3a/1b and 2a/1a) and three inter-subtype recombination types (1b/1a, 1a/1c and 4a/4d) have been described, based on analysis of either full-length or partial genome sequences [9]. Specifically, the 2k/1b recombinants have been demonstrated in Russia [16], Georgia, Estonia [17], Ireland [18], Uzbekistan [19], and Cyprus [20], and these are still circulating within Europe [21,22]. While it remains to be established, Morel et al. have suggested that genetic recombination may have important implications for HCV diagnosis, therapy, and epidemiology [21].

To date, only a few recombinants have been identified by analysis of a large number of complete genome sequences, and many recombination events have been identified by analyses of partial genome sequences [9]. It might be expected that, analysis

of partial genome sequences could underestimate both the true level of recombination in HCV and may not provide an accurate identification of the breakpoints involved [9,21]. In the present study, we have carried out an analysis of HCV recombination using the large number (n = 1278) of all available full length detailed genome sequences.

Datasets and Methods

1278 nucleotide sequences of HCV were downloaded from the Los Alamos HCV database (http://hcv.lanl.gov/content/index) on October 5th, 2011. The full length HCV genome is approximately 9600 bp in size. However, only the coding regions (approximately 9000 bp) are used in our analysis. In addition, a virus sequence of canine origin [23] was downloaded from GenBank and included as an outgroup in the phylogenetic analyses.

The DNA sequences were initially translated into protein sequences. The protein sequences were aligned using Clustal Omega [24] and the alignment was adjusted manually in Bioedit [25]. The DNA sequence alignment was then made using the protein alignment as a template. The DNA alignment was 9270 bp in length. We applied four strategies to subdivide the alignment into sub-datasets (Table S1). The first strategy subdivides the full-length genome alignment into 15 sub-datasets, with the first 14 sub-datasets 600 bp long and the last one 870 bp long. The second strategy cuts the alignment into 18 sub-datasets, with the first 17 sub-datasets 500 bp long and the last one 770 bp long. The third strategy splits the alignment into 23 sub-datasets, with the first 22 sub-datasets 400 bp long and the last one 470 bp long. The last strategy subdivides the alignment into 31 sub-datasets, with the first 30 sub-datasets 300 bp long and the last one 270 bp long. Phylogenetic analysis of the whole genome alignment and all of the sub-datasets was carried out using RAxML [26] under the GTRCAT approximation [27] and random starting trees. 1000, 600, 500, 400 and 300 rapid bootstrap replicates were performed for the full-length genome dataset, sub-datasets split using the first, second, third and fourth subdivision strategies, respectively. All other parameters were set to default. All of the trees are available on request from the authors. Trees were visualized using Dendroscope [28].

Information on these sequences, including genotype, subtype, and recombination, was downloaded from the database. This information was validated using the phylogenetic tree, constructed using the whole genome sequences, to correct potential genotype or subtype misclassifications and was used as background information. Subtyping information from each phylogenetic tree, constructed using the sub-datasets, was compared to the background information, on an individual basis. For each sequence, if all the information was concordant with the background information, this suggested the virus is not a recombinant. However, if a discrepancy between the background information and subtypes derived from the sub-datasets was identified, these sequences were analyzed further using multiple, independent computational methods described below.

Because the putative novel recombinants belonged to genotype 6, only the sequences of genotype 6 HCV (n = 77) were used for verification. All the potential novel putative recombinants were verified in a single run using the program RDP 3 [29]. The methods used included RDP [30], GENECONV [31], BootScan [32], Maxchi [33], Chimaera [34], SiSscan [35] and 3Seq [36]. The breakpoints were also defined by RDP. Similarity between the recombinants and their possible major and minor parents was estimated using Bioedit. BootScan, embedded in Simplot [37], was used to visualize the relationships among the recombinants and their possible parents, with a sequence (AF064490) from genotype 5 serving as an outgroup.

To further verify these recombination events, we extracted the NS5B genes of genotype 6 HCV from the whole alignment and split the alignment into two sub-alignments according to the breakpoints identified: the non-recombinant region and the recombinant region. We constructed phylogenetic trees using the non-recombinant NS5B gene regions and the recombinant regions, respectively. This was performed using PhyML [38]. To test the alternative topologies derived, we performed the Kishino-Hasegawa (KH) test [39] and calculated expected likelihood weights [40] using Tree-Puzzle [41].

In addition, to exclude the possibility that the detected recombination events are caused by lack of phylogenetic signals in the 3′-end of genotype 6 HCV, we used the likelihood mapping method [42], implemented in Tree-Puzzle, to test whether the datasets used for detecting recombination events are suitable for phylogenetic analysis. Three models (HKY [43], TN [44] and GTR [45]) were used, respectively. Similarly, only the NS5B genes of genotype 6 HCV were used in this analysis.

Table 1. Phylogenetic evidence of four novel inter-subtype recombination events.

GenBank No.	Country	Subtype	Phylogenetic evidence of recombination			
			Fragment length: 600 bp; number of fragments: 15	**Fragment length: 500 bp; number of fragments: 18**	**Fragment length: 400 bp; number of fragments: 23**	**Fragment length: 300 bp; number of fragments: 31**
EU246930	Viet Nam	6a	Fragment 1–14: 6a; fragment 15: 6o	Fragment 1–17: 6a; fragment 18: 6o	Fragment 1–21: 6a; fragment 22–23: 6o	Fragment 1–28: 6a; fragment 29: 6; fragment 30–31: 6o
EU246931	Viet Nam	6e	Fragment 1–14: 6e; fragment 15: 6h	Fragment 1–2: 6e; fragment 18: 6h	Fragment 1–21: 6e; fragment 22–23: 6h	Fragment 1–28: 6e; fragment 29: 6; fragment 30–31: 6h
EU246932	Viet Nam	6e	Fragment 1–14: 6e; fragment 15: 6o	Fragment 1–17: 6e; fragment 18: 6o	Fragment 1–21: 6e; fragment 22–23: 6o	Fragment 1–28: 6e; fragment 29: 6; fragment 30–31: 6o
EU246937	Thailand	6n	Fragment 1–14: 6n; fragment 15: 6o	Fragment 1–17: 6n; fragment 18: 6o	Fragment 1–21: 6n; fragment 22–23: 6o	Fragment 1–28: 6n; fragment 29: 6; fragment 30–31: 6o

doi:10.1371/journal.pone.0041997.t001

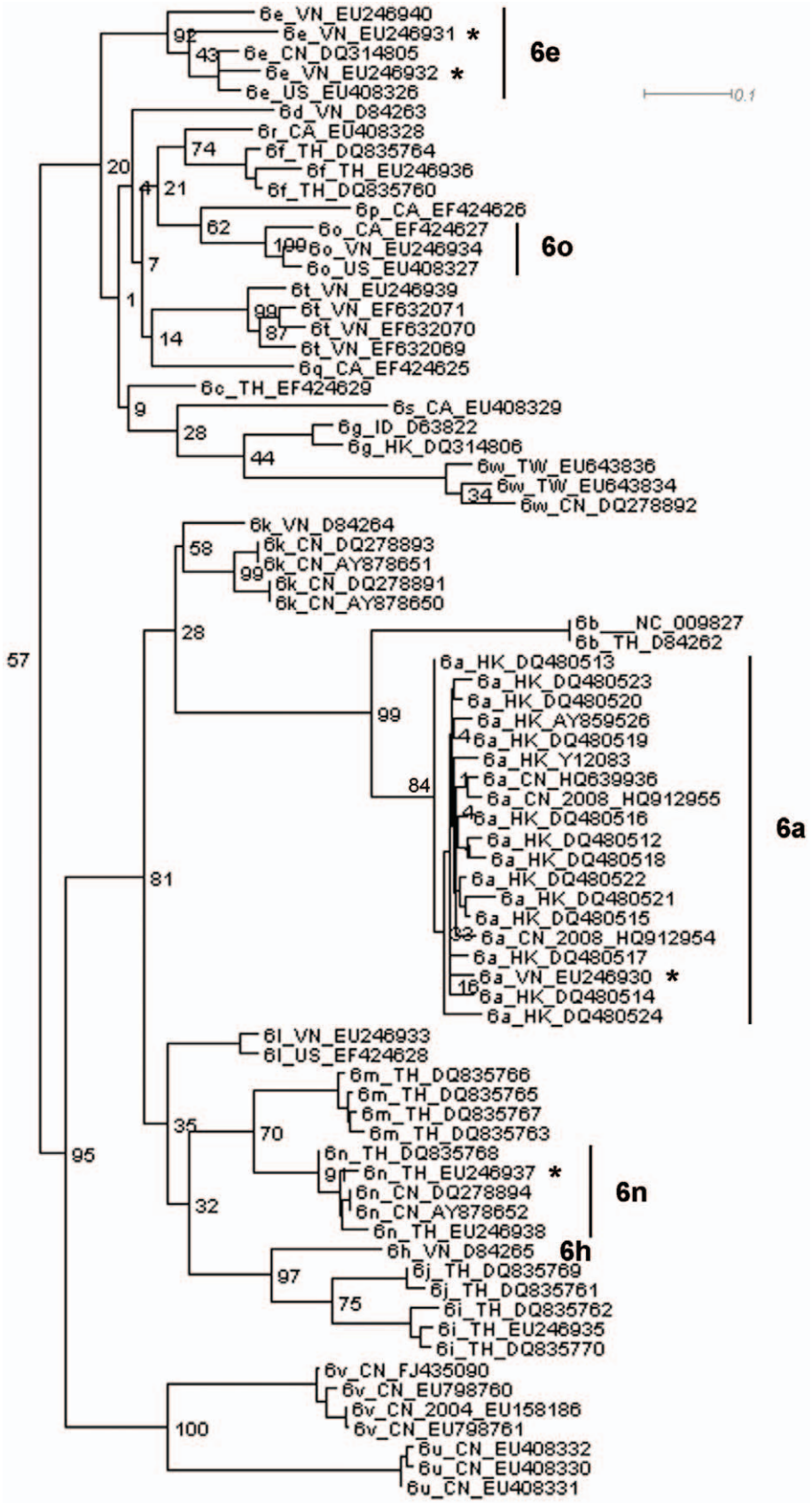

Figure 1. The genotype 6 part of the phylogenetic tree constructed using the first 600 bp of the alignment.
doi:10.1371/journal.pone.0041997.g001

Figure 2. The genotype 6 part of the phylogenetic tree constructed using the last 827 bp of the alignment.
doi:10.1371/journal.pone.0041997.g002

Table 2. Verification of the four novel inter-subtype recombinants by independent methods[a].

GenBank No.	Recombinant	RDP	GENECONV	BootScan	Maxchi	Chimaera	SiSscan	3Seq
EU246930	6a/6o	***	***	***	***	***	***	***
EU246931	6e/6h	***	***	***	***	**	***	***
EU246932	6e/6o	***	***	***	**	*	**	**
EU246937	6n/6o	***	***	***	**	***	**	**

[a]***means that the P value is smaller than 10^{-20} and **means that the P value is smaller than 10^{-10}.
*means that the P value is smaller than 10^{-5}.
doi:10.1371/journal.pone.0041997.t002

Results

Phylogenetic Analysis of the Full-length Genome Sequences

Phylogenetic analysis of the 1278 full-length genome sequences supports the current classification of HCV into seven genotypes, 1–7 (File S1). The number of sequences belonging to genotype 1 was 993, accounting for approximately 78% of the whole dataset, while that of genotype 2 was 116 (9%). Genotypes 3–7 included 33, 47, 5, 77 and 1 sequence, respectively.

Inter-genotype Recombination

By comparing phylogenetic signals from different subdivided fragments of the full-length genome sequences, we identified nine inter-genotype HCV recombinants. They belong to five recombination types, 2/5 (n = 2), 2b/6w (n = 1), 2b/1a (n = 1), 2b/1b (n = 1), and 2k/1b (n = 4), respectively (Table S2). All of these have been previously described [9]. No novel inter-genotype recombinants were found.

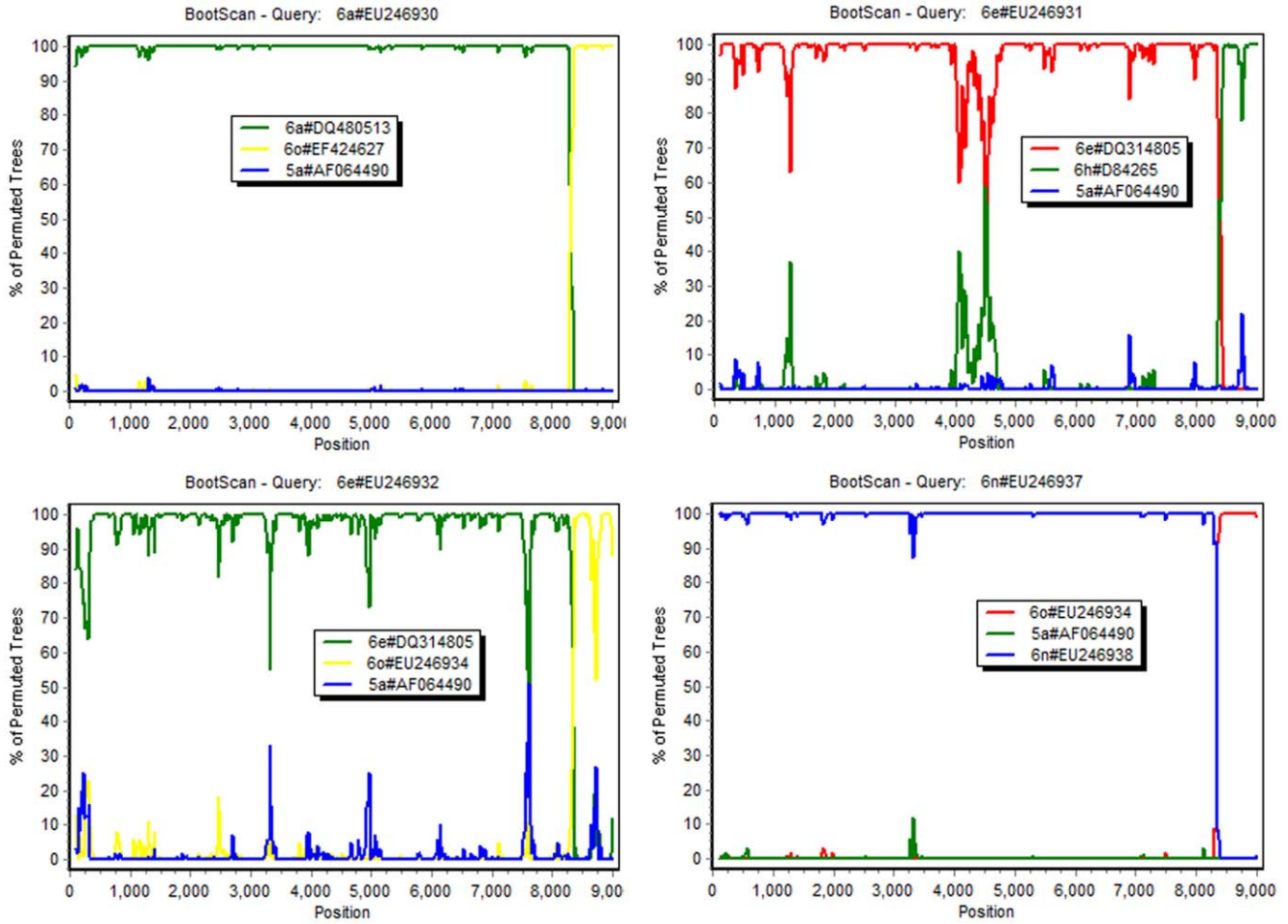

Figure 3. BootScan analysis of the novel recombinants and their possible parents. The coding region for the NS5B protein starts from position 7327 and ends at position 9099 in our alignment. Different colors represent sequences of different subtypes, two of which are the possible parental sequences and one of which (5a#AF06490) serves as an outgroup.
doi:10.1371/journal.pone.0041997.g003

Table 3. Breakpoints of the four novel inter-subtype recombinants[a].

Recombinant	In alignment		Length (bp)	Sequence similarity (%)	
	Begin	End		Major parent	Minor parent
EU246930 (6a)	8345	9073	729	DQ480513(6a): 77.3	EF424627(6o): 93.8
EU246931 (6e)	8356	9019	664	DQ314805(6e): 75.9	D84265(6h): 94.1
EU246932 (6e)	8358	8977	620	DQ314805(6e): 83.5	EU246934(6o): 99.3
EU246937 (6n)	8372	9033	662	EU246938(6n): 79.4	EU246934(6o): 96.8

[a]Sequence similarity between the recombinants and their major and minor parents is estimated only using the recombined regions. The major and minor parents of the recombinants are identified by RDP.

doi:10.1371/journal.pone.0041997.t003

Inter-subtype Recombination

Phylogenetic trees constructed using different sequence fragments can be used to find potential inter-subtype recombination events. In all, five inter-subtype recombinants were identified. The 1a/1c recombinant (AY651061) has already been reported [46] and was not further studied. The remaining four sequences, EU246930, EU246931, EU246932 and EU246937, are shown for the first time to be recombinants. These four sequences were isolated from Vietnam and Thailand and have been reported to belong to subtypes 6a, 6e, 6e and 6n, respectively [47]. Phylogenetic analysis of the full-length genome sequences confirmed this subtype classification (data not shown). However, phylogenetic trees estimated using the 600 bp (n = 15), 500 bp (n = 18), 400 bp (n = 23) and 300 bp (n = 31) fragments were consistent and demonstrated that EU246930, EU246931, EU246932 and EU246937 are 6a/6o, 6e/6h, 6e/6o, and 6n/6o recombinants, respectively (Table 1).

Figures 1 and 2 demonstrate how potential recombination events are identified from the trees. Figures 1 and 2 present the genotype 6 lineages of the phylogenetic trees constructed using the first fragment (600 bp in length) and the last fragment (827 bp in length) in the first sub-division strategy. In Figure 1, EU246930 (6a) is clustered within a lineage of 6a sequences and the bootstrap support value for this lineage is 84%. EU246931 and EU246932 (6e) fall within a cluster of 6e, with a bootstrap value of 92%, while EU246937 belongs to subtype 6n with a bootstrap value of 91%. However, in Figure 2, different phylogenetic relationships are found. EU246930 (6a), EU246932 (6e) and EU246937 (6n) are clustered with a lineage of subtype 6o sequences and the bootstrap support is 98%, while EU246931 (6e) forms a separate lineage with D84265 (6h) with 100% bootstrap support. Employing this approach, we analyzed all the trees and summarized the discordant phylogenetic signals suggesting evidence of recombination.

Further verification of these four recombinants was performed using RDP (Table 2). The four inter-subtype recombination events are supported by seven methods with significant p values (Table 2). The relationships within the recombinants with the potential major and minor parents identified by RDP were visualized using BootScan (Figure 3), which confirmed the recombination events.

Phylogenetic analyses and BootScan analysis indicate that the breakpoints of the four recombinants are located within the NS5B region (Table 1, Figure 3). Breakpoints of the four recombinants defined by RDP are consistent with the result obtained by phylogenetic analysis. However, the locations are not exactly the same in each case and the length of the recombined segments ranged from 620 bp for EU246932 to 729 bp for EU246930 (Table 3).

Results from the KH test and ELW were consistent and both of them supported that the phylogenies derived the non-recombinant region and the recombinant region were significantly different (Table 4).

Likelihood mapping analysis of NS5B gene sequences of genotype 6 HCV using different models were congruent. All of them showed that the tree-likeness of the NS5B gene was very high, with the sum of A_1, A_2, and A_3 ranging from 95.5% to 96.0% (Table 5). In contrast, the value of A_7, which is evidence to support the star-likeness, was relatively small, ranging from 1.2% to 1.8%. In particular, the recombinant regions (8358–9099) also displayed very high probability of tree-likeness (Table 5).

Discussion

Recombination in HCV has been considered a rare event. This is supported by the observation of superinfection exclusion, where an established virus infection prevents or interferes with subsequent infection by a second virus [10]. The first naturally occurring inter-genotype HCV recombinant was identified in 2002 [16]. This recombinant became established and is still circulating in some European countries [21,22]. So far, seven inter-genotype recombination types have been described [9]. Here, we identify nine inter-genotype recombinants and they belong to five inter-genotype recombination types, 2/5, 2b/6w, 2b/1a, 2b/1b and 2k/1b, respectively. All of these have been previously reported and no new or novel inter-genotype recombinants are found in this analysis.

Table 4. Likelihood mapping of NS5B gene sequences of genotype 6 HCV.

	HKY	TN	GTR
NS5B (7327–9099)	A_1+A_2+A_3: 95.6%; A_7: 1.8%	A_1+A_2+A_3: 95.5%; A_7: 1.8%	A_1+A_2+A_3: 96.0%; A_7: 1.2%
NS5B1 (7327–8357)	A_1+A_2+A_3: 90.1%; A_7: 5.3%	A_1+A_2+A_3: 90.1%; A_7: 5.4%	A_1+A_2+A_3: 91.6%; A_7: 3.5%
NS5B2 (8358–9099)	A_1+A_2+A_3: 91.1%; A_7: 4.9%	A_1+A_2+A_3: 90.3%; A_7: 5.6%	A_1+A_2+A_3: 93.2%; A_7: 2.7%

doi:10.1371/journal.pone.0041997.t004

Table 5. Statistical tests for alternative topologies derived from the recombination events detected*.

Event	Recombinant	Trees derived from (nt)		Test 1 vs 2			Test 2 vs 1		
		Dataset1	Dataset2	-lk best	p-SH	c-ELW	-lk best	p-SH	c-ELW
EU246930	8345–9073	7327–8344	8245–9073	18837.90	$<10^{-4}$	$<10^{-4}$	12008.24	$<10^{-4}$	$<10^{-4}$
EU246931	8356–9019	7327–8355	8356–9019	20068.04	$<10^{-4}$	$<10^{-4}$	10942.59	$<10^{-4}$	$<10^{-4}$
EU246932	8358–8977	7327–8357	8358–8977	20495.72	$<10^{-4}$	$<10^{-4}$	10339.85	$<10^{-4}$	$<10^{-4}$
EU246937	8372–9033	7327–8371	8372–9033	19985.97	$<10^{-4}$	$<10^{-4}$	10887.61	$<10^{-4}$	$<10^{-4}$

*Only the NS5B gene regions were used for this analysis.
doi:10.1371/journal.pone.0041997.t005

So far, only one subtype of genotype 5, 5a, has been identified. The breakpoint of the 2/5 recombinants is identified to be at or near the NS2/NS3 junction, between residues 3420 and 3440 [48]. Our results confirm this finding. However, the sequence divergence between the 2/5 recombinants and 5a from position 3421 to the end of the genome is 34.5% (1%, standard deviation), which is higher than the 20% cutoff used to define a subtype. Therefore, it is likely that the 2/5 recombinants are derived from a putative subtype of genotype 5, rather than 5a. Further collecting and sequencing more HCV samples of genotype 5 is needed to reveal the real phylogenetic diversity of HCV and to trace the most likely parents of the 2/5 recombinants.

In our work, five inter-subtype recombinants were found through large-scale phylogenetic analyses. The 1a/1c recombinant sequence was identified in India and has already been reported [46]. However, the remaining four recombinants are described here for the first time. These recombination events were well supported by various recombination detection methods and were shown not to result from the lack of phylogenetic signal in the 3′-end of HCV genomes. Specifically, they represent four novel inter-subtype recombination types, 6a/6o, 6e/6o, 6e/6h and 6n/6o, respectively.

Although only a few HCV recombinants have been described, current evidence suggests that the NS2/NS3 junction may be a hotspot for HCV recombination [9]. However, a breakpoint has also been identified within NS5B [49] and this is mapped to position 8046 in our alignment which is different from the breakpoints identified in our study (8245, 8356, 8358 and 8372, respectively). Notably, while the breakpoints of the recombinants identified in our study are not identical, they are very close. At present, it is impossible to determine whether these recombinants have arisen from single or multiple recombination events.

Two previous studies have also shown that recombination can happen within a single subtype or a patient [50,51]. Sentandreu et al. analyzed 17712 sequences from 136 serum samples derived from 111 patients and found approximately 11% of the samples were potential recombinant sequences [51]. On this basis, they concluded that recombination should be considered as a potentially important molecular mechanism for HCV to generate novel genetic variants. However, because our dataset has approximately 1300 sequences, it is extremely difficult to study detailed phylogenetic relationships for each sequence within a subtype using our approach and therefore we did not investigate intra-subtype recombination.

The subdivision of the whole HBV genome into numerous sub-datasets has previously termed "fragment typing" and has been used to identify putative HBV recombinants [52]. We have also recently used a similar approach to detect HBV recombination [7]. In this work, we applied four strategies to split the whole genome alignment into sub-datasets of different lengths, with different start and end points. The results obtained from the four strategies are broadly in agreement. Therefore, we consider our approach to be very robust for the detection of inter-genotype and inter-subtype HCV recombinants and this is particularly useful when large datasets with thousands of genome sequences are involved. However, this has two limitations. First, it may not be effective for the detection of small recombined fragments of less than 100 bp, because the shorter the alignment is, the lower the power and sensitivity of the phylogenetic analysis. Second, it is difficult to detect intra-subtype recombination using this method. For some subtypes, such as 1a and 1b where there are a few hundred sequences available, it is difficult to detect the in-congruent phylogenetic signals by "eyeballing" the trees. In these cases, methods that are able to automatically detect potential recombination events, such as RDP, as used in this study, should be employed.

Previous computer simulation studies and empirical data have shown that different recombination detection methods have distinct features and no single method is best for all situations [34,53]. In this work, seven methods used for verification of the results obtained from phylogenetic analyses. These methods are based on different rationales and have been classified into different classes [34,53]. For example, RDP, BootScan and SiSscan are phylogeny-based, while GENECONV, Maxchi and Chimaera are substitution-based. Because all of these methods detected the four sequences as recombinants, this provides very convincing evidence that these recombinants have been properly designated.

In conclusion, we have performed a large scale phylogenetic analysis of 1278 full-length genome sequences to detect putative inter-genotype and inter-subtype recombinants. No new or novel inter-genotype recombinants were found. However, we have identified for the first time four novel inter-subtype recombinants. Our studies suggest that HCV recombination and its implications for both pathogenesis and clinical outcomes certainly warrant further study.

Supporting Information

Table S1 Four strategies to subdivide the alignment into sub-datasets.
(XLSX)

Table S2 List of the identified inter-genotype HCV recombinants.
(DOCX)

File S1 Phylogenetic analysis of 1278 complete HCV genome sequences.
(TXT)

Acknowledgments

We thank Dr. XY LANG, Dr. XN WANG and their colleagues in the Supercomputing Center, Computer Network Information Center of The Chinese Academy of Sciences for their help in installing and optimizing RAxML on the SCIGRID.

Author Contributions

Conceived and designed the experiments: WS. Performed the experiments: WS ITF CZ WZ. Analyzed the data: WS ITF. Wrote the paper: WS WWH DGH.

References

1. Choo QL, Kuo G, Weiner AJ, Overby LR, Bradley DW, et al. (1989) Isolation of a cDNA clone derived from a blood-borne non-A, non-B viral hepatitis genome. Science 244: 359–362.
2. Simmonds P (2004) Genetic diversity and evolution of hepatitis C virus–15 years on. J Gen Virol 85: 3173–3188.
3. Dubuisson J (2007) Hepatitis C virus proteins. World J Gastroentero 13: 2406–2415.
4. Smith DB, Pathirana S, Davidson F, Lawlor E, Power J, et al. (1997) The origin of hepatitis C virus genotypes. J Gen Virol 78: 321–328.
5. Burke DS (1997) Recombination in HIV: an important viral evolutionary strategy. Emerg Infect Dis 3: 253–259.
6. Simmonds P, Midgley S (2005) Recombination in the genesis and evolution of hepatitis b virus genotypes. J Virol 79: 15467–15476.
7. Shi W, Carr MJ, Dunford L, Zhu C, Hall WW, et al. (2012) Identification of Novel Inter-genotypic Recombinants of Human Hepatitis B Viruses by Large-scale Phylogenetic analysis. Virology 427: 51–59.
8. Yun Z, Lara C, Johansson B, Lorenzana de Rivera I, Sonnerborg A (1996) Discrepancy of hepatitis C virus genotypes as determined by phylogenetic analysis of partial NS5 and core sequences. J Med Virol 49: 155–160.
9. Gonzalez-Candelas F, Lopez-Labrador FX, Bracho MA (2011) Recombination in hepatitis C virus. Viruses 3: 2006–2024.
10. Tscherne DM, Evans MJ, von Hahn T, Jones CT, Stamataki Z, et al. (2007) Superinfection exclusion in cells infected with hepatitis C virus. J Virol 81: 3693–3703.
11. Matsubara T, Sumazaki R, Shin K, Nagai Y, Takita H (1996) Genotyping of hepatitis C virus: coinfection by multiple genotypes detected in children with chronic posttransfusion hepatitis C. J Pediatr Gastr Nutr 22: 79–84.
12. Toyoda H, Fukuda Y, Hayakawa T, Takayama T, Kumada T, et al. (1998) Characteristics of patients with chronic infection due to hepatitis C virus of mixed subtype: prevalence, viral RNA concentrations, and response to interferon therapy. Clin Infect Dis 26: 440–445.
13. Giannini C, Giannelli F, Monti M, Careccia G, Marrocchi ME, et al. (1999) Prevalence of mixed infection by different hepatitis C virus genotypes in patients with hepatitis C virus-related chronic liver disease. J Lab Clin Med 134: 68–73.
14. Asselah T, Vidaud D, Doloy A, Boyer N, Martinot M, et al. (2003) Second infection with a different hepatitis C virus genotype in a intravenous drug user during interferon therapy. Gut 52: 900–902.
15. Schijman A, Colina R, Mukomolov S, Kalinina O, Garcia L, et al. (2004) Comparison of hepatitis C viral loads in patients with or without coinfection with different genotypes. Clin Diagn Lab Immun 11: 433–435.
16. Kalinina O, Norder H, Mukomolov S, Magnius LO (2002) A natural intergenotypic recombinant of hepatitis C virus identified in St. Petersburg. J Virol 76: 4034–4043.
17. Tallo T, Norder H, Tefanova V, Krispin T, Schmidt J, et al. (2007) Genetic characterization of hepatitis C virus strains in Estonia: fluctuations in the predominating subtype with time. J Med Virol 79: 374–382.
18. Moreau I, Hegarty S, Levis J, Sheehy P, Crosbie O, et al. (2006) Serendipitous identification of natural intergenotypic recombinants of hepatitis C in Ireland. Virol J 3: 95.
19. Kurbanov F, Tanaka Y, Avazova D, Khan A, Sugauchi F, et al. (2008) Detection of hepatitis C virus natural recombinant RF1_2k/1b strain among intravenous drug users in Uzbekistan. Hepatol Res 38: 457–464.
20. Demetriou VL, Kyriakou E, Kostrikis LG (2011) Near-full genome characterisation of two natural intergenotypic 2k/1b recombinant Hepatitis C virus isolates. Adv Virol 2011: 710438.
21. Morel V, Fournier C, Francois C, Brochot E, Helle F, et al. (2011) Genetic recombination of the hepatitis C virus: clinical implications. J Viral Hepatitis 18: 77–83.
22. Raghwani J, Thomas XV, Koekkoek SM, Schinkel J, Molenkamp R, et al. (2011) The origin and evolution of the unique HCV circulating recombinant form 2k/1b. J Virol 86: 2212–2220.
23. Kapoor A, Simmonds P, Gerold G, Qaisar N, Jain K, et al. (2011) Characterization of a canine homolog of hepatitis C virus. P Natl Acad Sci USA 108: 11608–11613.
24. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, et al. (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol Syst Biol 7: 539.
25. Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucl Acids Symp Ser 41: 95–98.
26. Stamatakis A, Ludwig T, Meier H. (2005) Raxml-iii: A fast program for maximum likelihood-based inference of large phylogenetic trees. Bioinformatics 21: 456–463.
27. Stamatakis A (2006) Phylogenetic models of rate heterogeneity: A high performance computing perspective, Proceedings of 20th IEEE/ACM International Parallel and Distributed Processing Symposium (IPDPS2006), High Performance Computational Biology Workshop, Rhodos, Greece.
28. Huson DH, Richter DC, Rausch C, Dezulian T, Franz M, et al. (2007) Dendroscope: An interactive viewer for large phylogenetic trees. BMC Bioinformatics 8(1): 460.
29. Martin DP, Lemey P, Lott M, Moulton V, Posada D, et al. (2010) Rdp3: A flexible and fast computer program for analyzing recombination. Bioinformatics 26: 2462–2463.
30. Martin D, Rybicki E (2000) RDP: detection of recombination amongst aligned sequences. Bioinformatics 16: 562–563.
31. Padidam M, Sawyer S, Fauquet CM (1999) Possible emergence of new geminiviruses by frequent recombination. Virology 265: 218–225.
32. Martin DP, Posada D, Crandall KA, Williamson C (2005) A modified bootscan algorithm for automated identification of recombinant sequences and recombination breakpoints. AIDS Res Hum Retroviruses 21: 98–102.
33. Maynard Smith J (1992) Analyzing the mosaic structure of genes. J Mol Evol 34: 126–129.
34. Posada D, Crandall KA (2001) Evaluation of methods for detecting recombination from DNA sequences: Computer simulations. P Natl Acad Sci USA 98: 13757–13762.
35. Gibbs MJ, Armstrong JS, Gibbs AJ (2000) Sister-Scanning: a Monte Carlo procedure for assessing signals in recombinant sequences. Bioinformatics 16: 573–582.
36. Boni MF, Posada D, Feldman MW (2007) An exact nonparametric method for inferring mosaic structure in sequence triplets. Genetics 176: 1035–1047.
37. Lole KS, Bollinger RC, Paranjape RS, Gadkari D, Kulkarni SS, et al. (1999) Full-length human immunodeficiency virus type 1 genomes from subtype c-infected seroconverters in india, with evidence of intersubtype recombination. J Virol 73: 152–160.
38. Guindon S, Gascuel O (2003) A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst Biol 52(5): 696–704.
39. Kishino H, Hasegawa M (1989) Evaluation of the maximum likelihood estimate of the evolutionary tree topologies from DNA sequence data, and the branching order in Hominoidea. J Mol Evol 29: 170–179.
40. Strimmer K, Rambaut A (2002) Inferring confidence sets of possibly misspecified gene trees. Proc R Soc Lond B 269: 137–142.
41. Schmidt HA, Strimmer K, Vingron M, von Haeseler A (2002) TREE-PUZZLE: maximum likelihood phylogenetic analysis using quartets and parallel computing. Bioinformatics 18: 502–504.
42. Strimmer K, von Haeseler A (1997) Likelihood-mapping: A simple method to visualize phylogenetic content of a sequence alignment. Proc Natl Acad Sci USA 94: 6815–6819.
43. Hasegawa M, Kishino H, Yano T (1985) Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. J Mol Evol 22(2): 160–174.
44. Tamura K, Nei M (1993) Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. Mol Biol Evol 10(3): 512–526.
45. Tavaré S (1986) Some Probabilistic and Statistical Problems in the Analysis of DNA Sequences. Lectures on Mathematics in the Life Sciences (American Mathematical Society) 17: 57–86.
46. Ross RS, Verbeeck J, Viazov S, Lemey P, Van Ranst M, et al. (2008) Evidence for a complex mosaic genome pattern in a full-length hepatitis C virus sequence. Evol Bioinform 4: 249–254.
47. Nopporpanth S, Poovorawan Y, Lien TX, Smits SL, Osterhaus AD, et al. (2008) Complete genome analysis of hepatitis C virus subtypes 6t and 6u. J Gen Virol 89: 1276–1281.
48. Legrand-Abravanel F, Claudinon J, Nicot F, Dubois M, Chapuy-Regaud S, et al. (2007) New natural intergenotypic (2/5) recombinant of hepatitis C virus. J Virol 81: 4357–4362.
49. Colina R, Casane D, Vasquez S, Garcia-Aguirre L, Chunga A, et al. (2004) Evidence of intratypic recombination in natural populations of hepatitis C virus. J Gen Virol 85: 31–37.
50. Moreno MP, Casane D, Lopez L, Cristina J (2006) Evidence of recombination in quasispecies populations of a Hepatitis C Virus patient undergoing anti-viral therapy. Virol J 3: 87.
51. Sentandreu V, Jimenez-Hernandez N, Torres-Puente M, Bracho MA, Valero A, et al. (2008) Evidence of recombination in intrapatient populations of hepatitis C virus. PloS One 3: e3239.
52. Yang J, Xing K, Deng R, Wang J, Wang X (2006) Identification of hepatitis b virus putative intergenotype recombinants by using fragment typing. J Gen Virol 87: 2203–2215.
53. Posada D (2002) Evaluation of methods for detecting recombination from DNA sequences: Empirical data. Mol Biol Evol 19: 708–717.