Published in final edited form as:

Virus Res. 2010 February ; 147(2): 202–207. doi:10.1016/j.virusres.2009.11.001.

Conflicting selection pressures target the NS3 protein in hepatitis C virus genotypes 1a and 1b

Stephanie Jiménez Irausquin and Austin L. Hughes*

Department of Biological Sciences, University of South Carolina, Columbia, SC 29208, USA

Abstract

Analysis of complete polyprotein-encoding sequences of the two most prevalent genotypes of hepatitis C virus (HCV-1a and HCV-1b) revealed evidence of abundant, slightly deleterious nonsynonymous variants subject to ongoing purifying selection. In the case of both HCV-1a and HCV-1b, the NS3 protein demonstrated a high incidence of forward-and-backward or parallel nonsynonymous changes in CTL epitopes as measured by the phylogenetic consistency index. These results imply that certain non-synonymous mutations have occurred frequently throughout the HCV-1a and HCV-1b phylogenies in the codons encoding the epitopes in NS3. This pattern is best explained by the frequent re-occurrence of the same set of escape mutations in CTL epitopes of NS3, which are selectively favored within hosts presenting the class I major histocompatibility complex molecule, but subject to purifying selection in the population at large. This pattern was more pronounced in HCV-1b than in HCV-1a, suggesting that there may be differences between the two genotypes with respect to NS3's interaction with host immune recognition.

Keywords

CD8+ T-lymphocytes; Hepatitis C virus; Natural selection; Sequence polymorphism

1. Introduction

Hepatitis C virus (HCV) is a positive-sense RNA virus belonging to the *Flaviviridae* family and is estimated to infect 130 million people globally (Alter, 2007). Although 30% of those infected with HCV are able to resolve the virus spontaneously, the remaining 70% develop persistent HCV infection and are threatened with the risk of developing liver cirrhosis and hepatocellular carcinoma (Lauer and Walker, 2001; Muller, 1996; Timm and Roggendorf, 2007). A high mutation rate, which results from both a high replication rate and a virally encoded RNA polymerase that lacks a proof reading function, contributes to the highly polymorphic nature of HCV (Neumann et al., 1998). This sequence variation is made apparent by phylogenetic analyzes that have identified six major genotypes, which have been further divided into several distinct subtypes (Simmonds et al., 1994, 2005; Timm and Roggendorf, 2007). The most prevalent genotypes include HCV-1a and HCV-1b (Kuiken et al., 2005), with HCV-1a being widely distributed in Northern Europe and the United States and HCV-1b being the most common worldwide (Simmonds et al., 2005).

© 2009 Elsevier B.V. All rights reserved.

*Corresponding author at: Department of Biological Sciences, Coker Life Sciences Building, University of South Carolina, 715 Sumter Street, Columbia, SC 29208, USA. Tel.: +1 803 777 9186. Austin@biol.sc.edu (A.L. Hughes).

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.virusres.2009.11.001.

Several studies have suggested escape from recognition by the host class I major histocompatibility complex (MHC) and CD8⁺ cytotoxic T cell (CTL) system as a mechanism for HCV persistence. For example, experimental infection of chimpanzees (Erickson et al., 2001; Weiner et al., 1995) as well as analyzes of viral sequences obtained from human patients (Cox et al., 2005; Gaudieri et al., 2006; Guglietta et al., 2005; Guo et al., 2004; Neuman-Haefelin et al., 2008; Poon et al., 2007; Ray et al., 2005; Seifert et al., 2004; Tester et al., 2005; Timm et al., 2004, 2007), have provided evidence of selection favoring escape mutations in CTL epitopes of HCV. Despite evidence that escape mutants in CTL epitopes may be selectively favored within infected hosts, there is also evidence that CTL epitopes demonstrate strong purifying selection at the population level; i.e., selection acting to eliminate most nonsynonymous mutations (Hughes et al., 2007; Irausquin and Hughes, 2008).

These contradictory findings may be explained by the hypothesis that though CTL escape mutants may prove beneficial to the virus with regard to immune evasion, they may also impose a cost to the virus in terms of decreased viability or transmissibility. Several studies have made observations that confer with this hypothesis. For example, a number of studies have noted that CTL epitopes at which escape mutations had previously occurred may revert to the population consensus sequence in the absence of the presenting class I MHC allele (Kuntzen et al., 2007; Ray et al., 2005; Timm et al., 2004). While another study found an evidence that certain CTL mutants, such as the mutants in the NS3 1073–81 CTL epitope, are harmful to the virus and can diminish protease activity and RNA replication (Söderholm et al., 2006). In addition, Irausquin and Hughes (2008) found evidence of an evolutionary conflict between immune escape and functional constraint in the evolution of the NS3 protein of HCV-1b. The CTL epitopes in NS3 showed evidence of numerous forward-and-backward changes at nonsynonymous sites, just as predicted if certain amino acid replacements were favored within hosts expressing the presenting class I MHC molecule but disfavored in other hosts (Irausquin and Hughes, 2008).

Here, we analyze coding sequence polymorphism in complete polyprotein-encoding sequences of both HCV-1a and HCV-1b, in order to test for evidence of conflicting evolutionary pressures on CTL epitopes of both genotypes. We test for evidence of both (1) a reduction in the effectiveness of purifying selection, and (2) forward-and-backward or parallel changes at nonsynonymous sites. These analyzes enable us to predict differences between HCV-1a and HCV-1b with respect to the major focus of CTL evasion within the human host.

2. Methods

2.1. Sequences analyzed and phylogenetic analysis

We analyzed 366 complete polyprotein-encoding nucleotide sequences of HCV-1a and 305 complete polyprotein-encoding nucleotide sequences of HCV-1b (for GenBank accession numbers see Supplementary Table S1) from the Hepatitis C Virus Database (<http://www.hcvdb.org/index.asp>). Sequences were subsequently aligned at the amino acid level using the ClustalW program (Larkin et al., 2007); this alignment was then imposed on the nucleotide sequences. Sequences containing undetermined nucleotides, premature stop codons and/or gaps were excluded from the analysis; recombinant sequences and sequences from non-human hosts were also excluded from the analysis.

Using the PAUP* program, version 4.0b10 (Swofford, 2003), we constructed a neighbor-joining (NJ) tree (Saitou and Nei, 1987) of HCV-1a and HCV-1b sequences, based on the proportion of nucleotide distance (Supplementary Fig. S1). The bootstrap method was used to assess the reliability of internal branches in the phylogenetic tree (Felsenstein, 1985);

1000 bootstrap replicates were used (Supplementary Fig. S1). We then used the maximum-parsimony method (Swofford, 2003) to reconstruct the last common ancestor for both HCV-1a and HCV-1b sequences.

On the basis of the NJ trees, we computed the consistency index (CI; Swofford, 2003) for each polymorphic nonsynonymous site based on the NJ tree. For a given site, CI is defined as the ratio of the minimum possible number of changes at a site to the number of changes at that site inferred assuming a given tree. CI is thus inversely related to the degree of homoplasy (forward-and-backward or parallel evolutionary changes) occurring at a site assuming a given tree. In these analyses, we excluded singleton sites (i.e., sites at which a difference from the most abundant nucleotide was found in only one of the sequences analyzed) because CI at such sites is trivial. Note also that our use of CI does not depend on the assumption that the NJ trees reconstructed phylogenetic relationships with 100% accuracy. Rather, we simply use CI as a measure of the extent to which a given site provides a phylogenetic signal consistent with the majority of sites in the data matrix (Irausquin and Hughes, 2008). It might be expected that CI would be somewhat higher in a larger sample of sequences because increasing the number of sequences sampled from a viral population increases the incidence of rare variants (Hughes, 2009), some of which may show homoplasy. Therefore, we did not compare CI directly between the two genotypes, since the numbers of sequences analyzed differed between genotypes.

2.2. Nucleotide sequence analysis

We estimated gene diversity at individual polymorphic synonymous and nonsynonymous nucleotide sites by the formula:

$$1 - \sum_{i=1}^n x_i^2$$

where n is the number of alleles and x_i is the frequency of the i th allele (Nei, 1987, p. 177). We classified single nucleotide polymorphisms as either synonymous or nonsynonymous based on their effect of the encoded amino acid sequence. Ambiguous sites, sites at which both synonymous and nonsynonymous variants occurred or at which the polymorphism could be considered synonymous or nonsynonymous depending on the pathway taken by evolution, were excluded. In HCV-1a, 4890 (54.4%) of 8988 sites were polymorphic; of these 2086 (42.7%) were synonymous, 1649 (33.7%) were nonsynonymous, and 1155 (23.6%) were ambiguous. In HCV-1b, 5200 (57.9%) of 8988 sites were polymorphic sites; of these 1950 (37.5%) were synonymous, 1902 (36.6%) were nonsynonymous, and 1348 (25.9%) were ambiguous.

We designated predicted epitopes presented by HLA class I MHC to CD8+ T cells as CTL epitopes. In comparing epitope and non-epitope regions, all epitopes for all proteins were defined by following the Los Alamos Hepatitis C Immunology Database (<http://hcv.lanl.gov/content/immuno/immuno-main.html>); only epitopes from human with known HLA presenting antigens were included (Yusim et al., 2005; Supplementary Table S2). Although most of these epitopes were originally discovered in HCV-1a, many were conserved in at least some HCV-1b sequences. Though CTL epitopes presented by different alleles and loci were combined for statistical purposes, it is important to note that several amino residues formed part of epitopes presented by more than one HLA allelic product and sometimes by products of alleles at different HLA loci (Supplementary Table S2). Also worth mentioning is that CTL epitopes sometimes overlapped epitopes presented by class II MHC molecules to CD4+ T cells (CD4 epitopes). We did not analyze CD4 epitopes in the

present paper because in a previous study CD4 epitopes showed no evidence of a distinctive evolutionary pattern (Irausquin and Hughes, 2008). Also in the present paper we did not test for different patterns among the different HLA class I alleles, because in a previous study we found no significant differences among CTL epitopes presented by different HLA class I alleles (Irausquin and Hughes, 2008).

Using the MEGA4 program (Tamura et al., 2007), the number of synonymous substitutions per synonymous site (d_S) and the number of nonsynonymous (amino acid-altering) substitutions per nonsynonymous site (d_N) were estimated by Nei and Gojobori's (1986) method. Estimation of these quantities by more complex methods (Li, 1993; Yang and Nielsen, 2000) yielded essentially identical results, as it is expected because the number of substitutions per site was generally low (Hughes and French, 2007); therefore, we used Nei and Gojobori's method because it has a lower error of estimation (Nei and Kumar, 2000). The synonymous nucleotide diversity (symbolized π_S) is defined as the mean of d_S for all pairwise comparisons among a set of sequences, while the nonsynonymous nucleotide diversity (symbolized π_N) is the mean of d_N for all pairwise comparisons among a set of sequences. We used robust methods of statistical analysis not depending on a specific model of sequence evolution (Hughes et al., 2006). We used non-parametric methods to analyze gene diversity and CI, since these quantities were not normally distributed (Hollander and Wolfe, 1973). So called codon-based methods of analysis were not used because they depend on several questionable assumptions, including the unwarranted assumption that the existence of one or more codons with $d_N > d_S$ implies positive selection (Hughes, 2007).

3. Results

3.1. Synonymous and nonsynonymous polymorphism

In each of the 10 proteins of both HCV-1a and HCV-1b, synonymous nucleotide diversity was significantly greater than nonsynonymous nucleotide diversity in both CTL epitope and non-epitope regions ($P < 0.001$ in every case; Supplementary Table S3). Thus, overall, purifying selection predominated in CTL epitopes as in the remainder of proteins. The ratio of nonsynonymous to synonymous polymorphic sites in the complete polyprotein of HCV-1a (1649:2086 or 0.79) differed significantly from the ratio of nonsynonymous to synonymous changes (406:915 or 0.44) reconstructed by maximum-parsimony as having been fixed in the HCV-1a ancestor ($\chi^2 = 72.8$, 1 d.f., $P < 0.001$). Likewise, the ratio of nonsynonymous to synonymous polymorphic sites in the complete polyprotein of HCV-1b (1902:1950 or 0.98) differed significantly to the ratio of nonsynonymous to synonymous changes (406:915 or .44) reconstructed by maximum-parsimony as having been fixed in the HCV-1b ancestor ($\chi^2 = 138.4$, 1 d.f., $P < 0.001$). Therefore, both HCV-1a and HCV-1b displayed a relative excess of nonsynonymous polymorphic sites in comparison to nonsynonymous divergent changes between genotypes.

We estimated gene diversity at individual polymorphic sites in order to test for ongoing purifying selection at these sites. In HCV-1a, median gene diversity for the complete polyprotein at synonymous polymorphic sites (0.113, $N = 2086$) was significantly greater than that at nonsynonymous polymorphic sites (0.016, $N = 1649$, Mann-Whitney test, $P < 0.001$). Similarly in HCV-1b, median gene diversity for the complete polyprotein at synonymous polymorphic sites (0.147, $N = 1950$) was significantly greater than that at nonsynonymous polymorphic sites (0.019, $N = 1902$, Mann-Whitney test, $P < 0.001$). Likewise, when individual proteins and CTL epitope and non-epitope regions were analyzed separately for both genotypes, median gene diversities at synonymous polymorphic sites exceeded those at nonsynonymous sites; and the differences were highly significant in the case of most of the proteins (Supplementary Fig. S2). These results support the hypothesis

that in both genotypes, there were abundant nonsynonymous variants subject to ongoing purifying selection.

3.2. Polymorphic sites in epitopes

As reported previously in a study involving a smaller number of HCV-1b sequences (Irausquin and Hughes, 2008), in NS3 of HCV-1b, median gene diversity at nonsynonymous polymorphic sites in CTL epitope regions (0.026) was significantly greater than that at nonsynonymous polymorphic sites in non-epitope regions (0.013; Fig. 1A). A similar pattern was seen in HCV-1a, where median gene diversity at nonsynonymous polymorphic sites in CTL epitope regions (0.027) was significantly greater than that at nonsynonymous polymorphic sites in non-epitope regions of NS3 (0.011; Fig. 1A). In nonsynonymous polymorphic sites in non-epitope regions of NS3, median gene diversity was significantly less in HCV-1a than in HCV-1b; but there was not a significant difference between HCV-1a and HCV-1b with respect to median gene diversity at nonsynonymous polymorphic sites in CTL epitope regions of NS3 (Fig. 1A). Besides NS3 only one other protein (NS4B) showed significantly higher median gene diversity at non-synonymous polymorphic sites in CTL epitopes than in non-epitope regions, but this pattern was only seen in HCV-1a (Supplementary Fig. S2).

In HCV-1a median gene diversity in non-epitope regions was significantly lower in NS3 (0.011) than in the other proteins (0.016; Fig. 1). Similarly in HCV-1b, median gene diversity in non-epitope regions was significantly lower in NS3 (0.013) than in the other proteins (0.019; Fig. 1). However, in CTL epitope regions of HCV-1b, median gene diversity was significantly higher in NS3 (0.026) than in other proteins (0.013; Fig. 1). This pattern was not seen in the case of CTL epitope regions of HCV-1a, where median gene diversity in NS3 (0.027) did not differ significantly from that in other proteins (0.022). Thus NS3 of HCV-1b showed a unique pattern of relaxed purifying selection on CTL epitopes.

HCV-1a and HCV-1b also showed different patterns of nonsynonymous polymorphism in proteins other than NS3. In HCV-1b, the median gene diversity at nonsynonymous polymorphic sites in CTL epitopes in proteins other than NS3 (0.013) was significantly less than that in non-epitope regions of the same proteins (0.019; Fig. 1B). However, in HCV-1a, the median gene diversity at non-synonymous polymorphic sites in CTL epitopes in proteins other than NS3 (0.022) was greater than that in non-epitope regions of the same proteins (0.016), although the difference was not significant (Fig. 1B). In nonsynonymous polymorphic sites of non-epitope regions in all proteins other than NS3, median gene diversity was significantly less in HCV-1a than in HCV-1b (Fig. 1B).

3.3. Patterns of homoplasy

In HCV-1b, median CI was significantly greater in non-epitope regions (0.333) than in CTL epitope regions (0.250) of NS3 (Fig. 2A). Similarly, in HCV-1a, median CI was significantly greater in non-epitope regions (0.250) than in CTL epitope regions (0.200) of NS3 (Fig. 2A). These results imply that forward-and-backward and parallel nonsynonymous changes occurred more frequently in CTL epitope regions than in non-epitope regions in NS3 of both genotypes.

In HCV-1b, proteins other than NS3 showed a pattern opposite of that seen in NS3; in the other proteins, median CI was significantly greater in CTL epitope regions (0.333) than in non-epitope regions (0.250) of the same proteins (Fig. 2B). In HCV-1a, on the other hand, median CI values in CTL epitope and non-epitope regions of proteins other than NS3 were identical (0.222; Fig. 2B). In the case of non-epitope regions of all proteins other than NS3, median CI was significantly greater in HCV-1b than HCV-1a (Fig 2B). Similarly, in CTL

epitope regions of all proteins other than NS3 median CI was significantly greater in HCV-1b than HCV-1a (Fig. 2B).

4. Discussion

Previous studies have shown evidence of purifying selection on protein-coding regions of HCV, including those that encode CTL epitopes (Hughes et al., 2007; Irausquin and Hughes, 2008). Here, analysis of complete polyprotein-encoding nucleotide sequences of both HCV-1a and HCV-1b revealed that gene diversity was significantly lower at nonsynonymous polymorphic sites than at synonymous polymorphic sites within both HCV-1a and HCV-1b populations. This finding provides evidence that in populations of these two HCV genotypes there are abundant slightly deleterious nonsynonymous variants subject to ongoing purifying selection.

In both HCV-1a and HCV-1b, CTL epitopes in the NS3 protein showed evidence of a relaxation of purifying selection. In NS3 of both genotypes, median gene diversity at nonsynonymous polymorphic sites in CTL epitope regions was elevated in comparison to non-epitope regions of the protein (Fig. 1A). However, the two genotypes also showed a striking difference with respect to the pattern of nonsynonymous polymorphism in NS3. In HCV-1b, median gene diversity in the CTL epitope regions of NS3 was unusually high in comparison to that in CTL epitope regions of other proteins (Fig. 1), while CI in the same epitopes was unusually low (Fig. 2A). The pattern of a relative excess of nonsynonymous polymorphic sites along with very low CI indicates the frequent re-occurrence of the same nonsynonymous mutations in the codons encoding these epitopes. On the other hand, in the case of HCV-1a, the CTL epitope regions of NS3 did not show a markedly different pattern from CTL epitopes in other proteins (Figs. 1 and 2).

These results suggest that the host CTL response may target NS3 to a greater degree in HCV-1b than in HCV-1a. Consistent with this hypothesis, results of several studies of HLA-A2 positive patients have suggested that epitopes from NS3 are more likely to be strongly immunodominant in HCV-1b than in other HCV genotypes (Chang et al., 1997; Lauer et al., 2002; Urbani et al., 2005a,b). In addition, there is evidence that NS3 differs functionally between HCV-1a and HCV-1b (Lam et al., 2003; Paredes and Blight, 2008), although it is unclear how these differences might relate to differences in the immune response.

A pattern of numerous forward-and-backward or parallel non-synonymous nucleotide changes in CTL epitopes, as observed most strikingly in NS3 of HCV-1b, can be explained by the occurrence of escape mutations that are selectively favored within hosts expressing the presenting MHC allele but are subject to purifying selection at the population level (Irausquin and Hughes, 2008). Empirical studies have provided evidence consistent with this interpretation in the case of several nonsynonymous variants showing low CI in NS3. For example, batches of anti-D immune globulin used in Ireland during 1977 and 1978 were contaminated with HCV-1b having the K → R mutation at position 1397 of CTL epitope NS3 1395–1403 (HSKKKCDEL; Kenny-Walsh, 1999; Ray et al., 2005; Salloum et al., 2008). The residue remains fixed in all HLA-B*08-positive subjects but reverted back to the original CTL epitope sequence in 50% of HLA-B*08-negative subjects several years after the transmission event (Ray et al., 2005; Salloum et al., 2008). Comparable findings were observed in an East German cohort also infected by contaminated anti-D immune globulins by way of a single-source outbreak of HCV genotype 1b in 1978 and 1979 (Salloum et al., 2008; Wiese et al., 2000). In both the HCV-1a and HCV-1b datasets used in our study, the nucleotide site corresponding to this amino acid polymorphism was a non-singular polymorphic site with low CI (0.100 in HCV-1a and 0.063 in HCV-1b). Moreover, there is evidence that, although CTL escape mutations that are observed in the HSKKKCDEL

epitope in vivo are found to be replication competent, the level of replication is reduced up to 60% compared to the HCV-1b subgenomic replicon Con1/ET (Salloum et al., 2008).

Another example, involving the Y → F replacement at position 1444 in the HLA-A*01-restricted CTL epitope NS3 1436–1444 (ATDALMTGY), was documented in a subject with acute HCV genotype 1a infection and was associated with reduced CTL response (Cox et al., 2005). Neuman-Haefelin et al. (2008) observed multiple reversals of the same mutation throughout a phylogeny of HCV-1b genotype sequences. In both of our datasets, the nucleotide site corresponding to this amino acid polymorphism was a non-singular polymorphic site with low CI (0.010 in HCV-1a and 0.019 in HCV-1b). Yet another study, involving the CTL epitope NS3 1073–1081 (CINGVCWTV) revealed two different amino acid replacements that caused reduced recognition by human HLA-A2-restricted CTLs (position 1074 I → V and position 1077 V → A), and the latter was also associated with reduced HLA-A2-binding avidity (Söderholm et al., 2006). Interestingly, this epitope more frequently carries a CINGCWTV sequence in genotype 1a and a CVNGVCWTV sequence in genotype 1b (Urbani et al., 2005a). In our datasets, the nucleotide site corresponding to the I → V polymorphism also showed low CI (0.125 in HCV-1a and 0.026 in HCV-1b). Similarly, the nucleotide site corresponding to the V → A polymorphism showed low CI (0.250 in HCV-1a and 0.077 in HCV-1b).

In conclusion, our results support the hypothesis that certain mutants in CTL epitopes of both HCV-1a and HCV-1b sequences are subjected to conflicting evolutionary pressures: positive selection favoring escape mutants within hosts expressing the presenting class I MHC molecule and purifying selection acting to remove them in the population at large. The fact that this pattern is more apparent in the NS3 protein of HCV-1b than in that of HCV-1a suggests that there are important differences in the cellular immune response to these two viral genotypes, just as there appear to be differences in the humoral immune response (Carlos et al., 2004).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This research was supported by Grant GM43940 from the National Institutes of Health to A.L.H.

References

- Alter M. Epidemiology of hepatitis C virus infection. *World J. Gastroenterol.* 2007; 13:2436–2441. [PubMed: 17552026]
- Carlos MP, Yamamura Y, Vu Q, Conzen K, Anderson DE, Torres JV. Humoral immunity to immunodominant epitopes of hepatitis C virus in individuals infected with genotypes 1a and 1b. *Clin. Immunol.* 2004; 111:22–27. [PubMed: 15093548]
- Chang K-M, Reherrmann B, McHutchinson JH, Pasquinelli C, Southwood S, Sette A. Immunological significance of cytotoxic T lymphocyte epitope variants in patients chronically infected by the hepatitis C virus. *J. Clin. Invest.* 1997; 100:2376–2385. [PubMed: 9410918]
- Cox AL, Mosbrugger T, Mao Q, Liu Z, Wang XH, Yang HC, Sidney J, Sette A, Pardoll D, Thomas DL, Ray SC. Cellular immune selection with hepatitis C virus persistence in humans. *J. Exp. Med.* 2005; 201:1741–1752. [PubMed: 15939790]
- Erickson AL, Kimura Y, Igarashi S, Eichelberger J, Houghton M, Sidney J, McKinney D, Sette A, Hughes AL, Walker CM. The outcome of hepatitis C virus infection is predicted by escape mutations in epitopes targeted by cytotoxic T lymphocytes. *Immunity.* 2001; 15:883–895. [PubMed: 11754811]

- Felsenstein J. Confidence limits on phylogenies: an approach using the bootstrap. *Evolution*. 1985; 39:783–791.
- Gaudieri S, Rauch A, Park LP, Freitas E, Herrmann S, Jeffrey G, Cheng W, Pfafferrott K, Naidoo K, Chapman R, Battegay M, Weber R, Telenti A, Furrer H, James I, Lucas M, Mallal SA. Evidence of viral adaptation to HLA class I-restricted immune pressure in chronic hepatitis C virus infection. *J. Virol.* 2006; 80:11094–11104. [PubMed: 17071929]
- Guglietta S, Garbuglia AR, Pacciani V, Scottà C, Perrone MP, Laurenti L, Spada E, Mele A, Capobianchi MR, Taliani G, Folgori A, Vitelli A, Ruggeri L, Nicosia A, Piccolella E, Del Porto P. Positive selection of cytotoxic T lymphocyte escape variants during acute hepatitis C virus infection. *Eur. J. Immunol.* 2005; 35:2627–2637. [PubMed: 16114108]
- Guo HZ, Yin Y, Wang WL, Zhang CS, Wang T, Wang Z, Zhang J, Cheng H, Wang HT. Sequence evolution of putative cytotoxic T cell epitopes in NS3 region of hepatitis C virus. *World J. Gastroenterol.* 2004; 10:847–851. [PubMed: 15040030]
- Hollander, M.; Wolfe, DA. *Nonparametric Statistical Methods*. New York: Wiley & Sons; 1973.
- Hughes AL. Looking for Darwin in all the wrong places: the misguided quest for positive selection at the nucleotide sequence level. *Heredity*. 2007; 99:364–373. [PubMed: 17622265]
- Hughes AL. Small effective population sizes and rare nonsynonymous variants in potyviruses. *Virology*. 2009; 393:127–134. [PubMed: 19695658]
- Hughes AL, French JO. Homologous recombination and the pattern of nucleotide substitution in *Ehrlichia ruminantium*. *Gene*. 2007; 387:31–37. [PubMed: 17005333]
- Hughes AL, Friedman R, Glenn NL. The future of data analysis in evolutionary genomics. *Curr. Genomics*. 2006; 7:227–234.
- Hughes AL, Hughes MA, Friedman R. Variable intensity of purifying selection on cytotoxic T-lymphocyte epitopes in hepatitis C virus. *Virus Res.* 2007; 123:147–153. [PubMed: 17005284]
- Irausquin SJ, Hughes AL. Distinctive pattern of sequence polymorphism in the NS3 protein of hepatitis C virus type 1b reflects conflicting evolutionary pressures. *J. Gen. Virol.* 2008; 89:1921–1929. [PubMed: 18632963]
- Kenny-Walsh E. Clinical outcomes after hepatitis C infection from contaminated anti-D immune globulin. Irish Hepatology Research Group. *N. Engl. J. Med.* 1999; 340:1228–1233. [PubMed: 10210705]
- Kuiken C, Yusim K, Boykin L, Richardson R. The Los Alamos HCV Sequence Database. *Bioinformatics*. 2005; 21:379–384. [PubMed: 15377502]
- Kuntzen T, Timm J, Berical A, Lewis-ximenez LL, Jones A, Nolan B, Schulze zur Wiesch J, Li B, Scheidewind A, Kim AY, Chung RT, Lauer GM, Allen TM. Viral sequence evolution in acute hepatitis C virus infection. *J. Virol.* 2007; 81:11658–11668. [PubMed: 17699568]
- Lam AM, Keeney D, Eckert PQ, Frick DN. Hepatitis C virus NS3 ATPases/helicases from different genotypes exhibit variations in enzymatic properties. *J. Virol.* 2003; 77:3950–3961. [PubMed: 12634355]
- Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG. ClustalW and ClustalX version 2.0. *Bioinformatics*. 2007; 23:2947–2948. [PubMed: 17846036]
- Lauer GM, Walker BD. Hepatitis C virus infection. *N. Engl. J. Med.* 2001; 345:41–52. [PubMed: 11439948]
- Lauer GM, Ouchi K, Chung RT, Nguyen TN, Day CL, Purkis DR, Reiser M, Kim AY, Lucas M, Klenerman P, Walker BD. Comprehensive analysis of CD8+ T-cell responses against hepatitis C virus reveals multiple unpredicted specificities. *J. Virol.* 2002; 76:6104–6113. [PubMed: 12021343]
- Li WH. Unbiased estimates of the rates of synonymous and nonsynonymous substitution. *J. Mol. Evol.* 1993; 36:96–99. [PubMed: 8433381]
- Muller R. The natural history of hepatitis C: clinical experiences. *J. Hepatol.* 1996; 24:52–54. [PubMed: 8836889]
- Nei, M. *Molecular Evolutionary Genetics*. New York: Columbia University Press; 1987.
- Nei M, Gojobori T. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol. Biol. Evol.* 1986; 3:418–426. [PubMed: 3444411]

- Nei, M.; Kumar, S. *Molecular Evolution and Phylogenetics*. New York: Oxford University Press; 2000.
- Neumann AU, Lam NP, Dahari H, Gretch DR, Wiley TE, Layden TJ, Perelson AS. Hepatitis C viral dynamics in vivo and the antiviral efficacy of interferon alpha therapy. *Science*. 1998; 282:103–107. [PubMed: 9756471]
- Neuman-Haefelin C, Frick DN, Wang JJ, Pybus OG, Salloum S, Narula GS, Eckart A, Biezyński A, Eiermann T, Klennerman P, Viazov S, Roggendorf M, Thimme R, Reiser M, Timm J. Analysis of evolutionary forces in an immunodominant CD8 epitope in hepatitis C virus at a population level. *J. Virol.* 2008; 82:3438–3451. [PubMed: 18216107]
- Paredes AM, Blight KJ. A genetic interaction between hepatitis C virus NS4B and NS3 is important for RNA replication. *J. Virol.* 2008; 82:10671–10683. [PubMed: 18715921]
- Poon AF, Kosakovsky Pond SL, Bennett P, Richman DD, Leigh Brown AJ, Frost SD. Adaptation to human populations is revealed by within-host polymorphisms in HIV-1 and hepatitis C virus. *PLoS Pathogens*. 2007; 3:e45. [PubMed: 17397261]
- Ray SC, Fanning L, Wang XH, Netski DM, Kenny-Walsh E, Thomas DL. Divergent and convergent evolution after a common-source outbreak of hepatitis C virus. *J. Exp. Med.* 2005; 201:1753–1759. [PubMed: 15939791]
- Saitou N, Nei M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* 1987; 4:406–425. [PubMed: 3447015]
- Salloum S, Oniangue-Ndza C, Neumann-Haefelin C, Hudson L, Giugliano S, aus dem Siepen M, Nattermann J, Spengler U, Lauer GM, Wiese M, Klennerman P, Bright H, Scherbaum N, Thimme R, Roggendorf M, Viazov S, Timm J. Escape from HLA-B*08-restricted CD8 T cells by hepatitis C virus is associated with fitness costs. *J. Virol.* 2008; 82:11803–11812. [PubMed: 18815309]
- Seifert U, Liermann H, Racanelli V, Halenius A, Wiese M, Wedemeyer H, Ruppert T, Rispeter K, Henklein P, Sijts A, Hengel H, Kloetzel PM, Rehmann B. Hepatitis C virus mutation affects proteasomal epitope processing. *J. Clin. Invest.* 2004; 114:250–259. [PubMed: 15254592]
- Simmonds P, Alberti A, Alter HJ, Bonino F, Bradley DW, Brechot C, Brouwer JT, Chan SW, Chayama K, Chen DS, et al. A proposed system for the nomenclature of hepatitis C viral genotypes. *Hepatology*. 1994; 19:1321–1324. [PubMed: 8175159]
- Simmonds P, Bukh J, Combet C, Deléage G, Enomoto N, Feinstone S, Halfon P, Inschaupé G, Kuiken C, Maertens G, Mizokami M, Murphy DG, Okamoto H, Pawlotsky JM, Penin F, Sablon E, Shin-I T, Stuyver LJ, Thiel HJ, Viazov S, Weiner AJ, Widell A. Consensus proposals for a unified system of nomenclature of hepatitis C virus genotypes. *Hepatology*. 2005; 42:962–973. [PubMed: 16149085]
- Söderholm J, Ahlén G, Kaul A, Frelin L, Alheim M, Barnfield C, Liljeström P, Weiland O, Milich DR, Bartenschlager R, Sällberg M. Relation between viral fitness and immune escape within the hepatitis C virus protease. *Gut*. 2006; 55:266–274. [PubMed: 16105887]
- Swofford, DL. *Phylogenetic Analysis Using Parsimony (*and Other Methods)*. Version 4. Sunderland, MA: Sinauer Associates; 2003. PAUP*.
- Tamura K, Dudley J, Nei M, Kumar S. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol. Biol. Evol.* 2007; 24:1596–1599. [PubMed: 17488738]
- Tester I, Smyk-Pearson S, Wang P, Wertheimer A, Yao E, Lewinsohn DM, Tavis JE, Rosen HR. Immune evasion versus recovery after acute hepatitis C virus infection from a shared source. *J. Exp. Med.* 2005; 201:1725–1731. [PubMed: 15939788]
- Timm J, Lauer GM, Kavanagh DG, Sheridan I, Kim AY, Lucas M, Pillay T, Ouchi K, Reyrol LL, Schulze zur Wiesch J, Gandhi RT, Chung RT, Bhardwaj N, Klennerman P, Walker BD, Allen TM. CD8 epitope escape and reversion in acute HCV infection. *J. Exp. Med.* 2004; 200:1593–1604. [PubMed: 15611288]
- Timm J, Li B, Daniels MG, Bhattacharya T, Reyrol LL, Allgaier R, Kuntzen T, Fischer W, Nolan BE, Duncan J, Schulze zur Wiesch J, Kim AY, Frahm N, Brander C, Chung RT, Lauer GM, Korber BT, Allen TM. Human leukocyte antigen-associated sequence polymorphisms in hepatitis C virus reveal reproducible immune responses and constraints on viral evolution. *Hepatology*. 2007; 46:339–349. [PubMed: 17559151]

- Timm J, Roggendorf M. Sequence diversity of hepatitis C virus: implications for immune control and therapy. *World J. Gastroenterol.* 2007; 13:4808–4817. [PubMed: 17828811]
- Urbani S, Amadei B, Cariani E, Fisicaro P, Orlandini A, Missale G, Ferrari C. The impairment of CD8 responses limits the selection of escape mutations in acute hepatitis C virus infection. *J. Immunol.* 2005a; 175:7519–7529. [PubMed: 16301660]
- Urbani S, Amadei B, Fisicaro P, Pilli M, Missale G, Bertolotti A, Ferrari C. Heterologous T cell immunity in severe hepatitis C virus infection. *J. Exp. Med.* 2005b; 201:675–680. [PubMed: 15753202]
- Weiner A, Erickson AL, Kansopon J, Crawford K, Muchmore E, Hughes AL, Houghton M, Walker CM. Persistent hepatitis C virus infection in a chimpanzee is associated with the emergence of a cytotoxic T lymphocyte escape variant. *Proc. Natl. Acad. Sci. U.S.A.* 1995; 92:2755–2759. [PubMed: 7708719]
- Wiese M, Berr F, Lafrenz M, Porst H, Oesen U. Low frequency of cirrhosis in a hepatitis C (genotype 1b) single-source outbreak in germany: a 20-year multicenter study. *Hepatology.* 2000; 32:91–96. [PubMed: 10869294]
- Yang Z, Nielsen R. Estimating synonymous and nonsynonymous substitution rates under realistic evolutionary models. *Mol. Biol. Evol.* 2000; 17:32–43. [PubMed: 10666704]
- Yusim K, Richardson R, Tao N, Szinger J, Funkhouser R, Korber B, Kuiken C. The Los Alamos Hepatitis C Immunology Database. *Appl. Bioinformatics.* 2005; 4:217–225. [PubMed: 16309340]

Fig. 1.

Median gene diversity at polymorphic nonsynonymous sites in CTL epitope (Ep) and non-CTL-epitope (NonEp) regions of (A) NS3 and (B) all other proteins. Numbers of sites in each category are shown. Mann–Whitney tests of the equality of median gene diversity between HCV-1a and HCV-1b sequences: ^a $P < 0.01$, ^b $P < 0.001$; between CTL epitope and non-CTL-epitope regions: ^c $P < 0.05$, ^d $P < 0.01$, ^e $P < 0.001$; and between NS3 and other proteins: ^f $P < 0.05$, ^g $P < 0.01$, ^h $P < 0.001$.

Fig. 2. Median CI at polymorphic nonsynonymous non-singleton sites in CTL epitope (Ep) and non-CTL-epitope (NonEp) regions of (A) NS3 and (B) all other proteins. Numbers of sites in each category are shown. Mann–Whitney tests of the equality of median CI between CTL epitope and non-CTL-epitope regions: ^a $P < 0.05$, ^b $P < 0.01$; and between NS3 and other proteins: ^c $P < 0.05$, ^d $P < 0.01$.