VIRUS EVOLUTION

Virus Evolution, 2022, **8(1)**, 1–14

DOI: https://doi.org/10.1093/ve/veac007

Advance access publication date: 2 February 2022

Research Article

Intra-host analysis of hepaciviral glycoprotein evolution reveals signatures associated with viral persistence and clearance

André Gömer,[1,2,†] Richard J. P. Brown,[3,‡] Stephanie Pfaender,[1] Katja Deterding,[4,5] Gábor Reuter,[6] Richard Orton,[7,§] Stefan Seitz,[8] C.- Thomas Bock,[9] Jessika M. V. Cavalleri,[10] Thomas Pietschmann,[11,12,13] Heiner Wedemeyer,[4,5] Eike Steinmann,[1] and Daniel Todt[1,11,14,*,∥]

[1]Department for Molecular and Medical Virology, Ruhr University Bochum, Universitätsstr. 150, Bochum 44801, Germany, [2]Institute of Virology, University of Veterinary Medicine Hannover, Foundation, Bünteweg 9, Hannover 30559, Germany, [3]Division of Veterinary Medicine, Paul Ehrlich Institute, Paul-Ehrlich-Straße 51-59, Langen 63225, Germany, [4]Department of Gastroenterology, Hepatology and Endocrinology, Hannover Medical School, Carl-Neuberg-Straße 1, Hannover 30625, Germany, [5]German Center for Infectious Disease Research (DZIF), HepNet Study-House, Hannover 30625, Germany, [6]Department of Medical Microbiology and Immunology, Medical School, University of Pécs, Szigeti út 12., Pécs 7624, Hungary, [7]MRC-University of Glasgow, Centre for Virus Research, Garscube Campus, 464 Bearsden Road, Glasgow G61 1QH, United Kingdom, [8]Department of Infectious Diseases, Molecular Virology, University of Heidelberg, Heidelberg 69120, Germany, [9]Division of Viral Gastroenteritis and Hepatitis Pathogens and Enteroviruses, Department of Infectious Diseases, Robert Koch Institute, Berlin 13353, Germany, [10]Clinical Unit of Equine Internal Medicine, University of Veterinary Medicine Vienna, Veterinärplatz 1, Vienna 1210, Austria, [11]Twincore, Centre for Experimental and Clinical Infection Research, Institute of Experimental Virology, Hannover 30625, Germany, [12]German Centre for Infection Research (DZIF), Partner Site Hannover-Braunschweig Site, Hannover 30625, Germany, [13]Cluster of Excellence RESIST (EXC 2155), Hannover Medical School, Hannover 30625, Germany and [14]European Virus Bioinformatics Centre (EVBC), Jena 07743, Germany

[†]https://orcid.org/0000-0002-4567-0441

[‡]https://orcid.org/0000-0002-3292-6671

[§]https://orcid.org/0000-0002-3389-4325

[∥]https://orcid.org/0000-0002-3564-1014

*Corresponding author: E-mail: daniel.todt@ruhr-uni-bochum.de

Abstract

Even 30 years after the discovery of the hepatitis C virus (HCV) in humans there is still no vaccine available. Reasons for this include the high mutation rate of HCV, which allows the virus to escape immune recognition and the absence of an immunocompetent animal model for vaccine development. Phylogenetically distinct hepaciviruses (genus *Hepacivirus*, family *Flaviviridae*) have been isolated from diverse species, each with a narrow host range: the equine hepacivirus (EqHV) is the closest known relative of HCV. In this study, we used amplicon-based deep-sequencing to investigate the viral intra-host population composition of the genomic regions encoding the surface glycoproteins E1 and E2. Patterns of E1E2 substitutional evolution were compared in longitudinally sampled EqHV-positive sera of naturally and experimentally infected horses and HCV-positive patients. Intra-host virus diversity was higher in chronically than in acutely infected horses, a pattern which was similar in the HCV-infected patients. However, overall glycoprotein variability was higher in HCV compared to EqHV. Additionally, selection pressure in HCV populations was higher, especially within the N-terminal region of E2, corresponding to the hypervariable region 1 (HVR1) in HCV. An alignment of glycoprotein sequences from diverse hepaciviruses identified the HVR1 as a unique characteristic of HCV: hepaciviruses from non-human species lack this region. Together, these data indicate that EqHV infection of horses could represent a powerful surrogate animal model to gain insights into hepaciviral evolution and HCVs HVR1-mediated immune evasion strategy.

Key words: Hepatitis C virus; equine hepacivirus; intra-host evolution; glycoprotein variability; hypervariable region.

1. Introduction

Thirty years ago, the hepatitis C virus (HCV) was discovered as the first member of the genus *Hepacivirus* (Simmonds et al. 2017) and thereafter characterized as a hepatotropic virus with high chronicity rates causing severe liver damage (Choo et al. 1989). Up until 2011, HCV-infected patients were treated with pegylated interferon-alpha (PEG-IFNα) combined with ribavirin (RBV), with limited sustained virologic responses (50–80 per cent, dependent on infecting genotype) and severe side effects (Manns, Wedemeyer, and Cornberg 2006). In 2011, PEG-IFNα/RBV therapy was replaced by the first generation Direct Acting Antivirals (DAAs), with cure rates above 90 per cent (Martinello et al. 2018). However, with 71 million chronically infected people worldwide, continued high-transmission rates, and no available vaccine, HCV

© The Author(s) 2022. Published by Oxford University Press.
This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (https://creativecommons.org/licenses/by-nc/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

Downloaded from https://academic.oup.com/ve/article/8/1/veac007/6520266 by Vetmeduni Vienna user on 07 March 2022

remains a serious global health burden, putting at risk the WHO's aim to eradicate HCV by 2030 (World Health Organization 2017). In 80 per cent of the cases, HCV leads to a chronic infection that often remains undetected as the early stages of the disease are asymptomatic. Subsequently, liver damage accumulates over years of chronic infection, potentially leading to liver fibrosis, cirrhosis, and hepatocellular carcinoma (Blum 2016). Additionally, even after HCV clearance which occurs only in a minority (20 per cent), patients do not develop protective immunity, allowing reinfection (Grebely et al. 2012; Bartenschlager et al. 2018). Furthermore, patients that had successfully been treated with DAAs still showed increased tumour recurrence rates and no complete reversal of immune signatures (Reig et al. 2016; Renzulli et al. 2018; Khera et al. 2021). HCV has developed multiple mechanisms to escape host-immune detection. One of these is its high replication capacity, paired with a high mutation rate caused by the low-fidelity polymerase resulting in the generation of a highly variable intra-host virus population (Martell et al. 1992; Bartenschlager and Lohmann 2000; Domingo, Sheldon, and Perales 2012). This enables the virus to quickly adapt to environmental changes, such as selective pressure mediated by the immune system or therapeutics (Vignuzzi et al. 2006; von Hahn et al. 2007; Joseph et al. 2015).

The E1E2 glycoproteins play a crucial role in HCV biology, as they must withstand high immune pressure elicited by targeted humoral responses to maintain persistent infection in the host. One mechanism by which HCV escapes E1E2 immune targeting is its high variability; however, the glycoproteins must also retain highly conserved regions such as receptor binding sites, which facilitate entry into host target cells. These conserved sites are protected by both a glycan shield and variable domains that act as an immunogenic decoy (Goffard and Dubuisson 2003; Lavie, Hanoulle, and Dubuisson 2018; Prentoe and Bukh 2018). The Hypervariable Region 1 (HVR1) is one of three hypervariable regions within the E1E2 protein and plays a central role in protecting the binding site for the entry receptor CD81. Furthermore, the glycoproteins enable HCV to associate with apolipoproteins derived from the host, which also provide the virion with an additional protective layer (Meunier et al. 2005; Fauvelle et al. 2016; Bankwitz et al. 2017).

For a prophylactic vaccine to effectively neutralize HCV and elicit protective immunity, it must overcome all these hurdles. HCV vaccine development is however hindered by the narrow species tropism of HCV, which infects only humans. The only immunocompetent animal model, the chimpanzee (*Pan troglodytes*), allows experimental infection but this is now ethically restricted (Bukh 2004; Burm et al. 2018).

Alternative animal models such as rodents are being used for HCV research, but do not support efficient HCV propagation, which makes them unsuitable for vaccine design. Attempts to genetically engineer mice to facilitate efficient HCV propagation are ongoing, but low replication rates are still a limitation (Dorner et al. 2013; Brown et al. 2020). Likewise, xenograft mouse models do not support high-level viral titers or are not immunocompetent. Moreover, adapting HCV to rodents was still not successful, lacking efficient replication and sufficient viral titers. Therefore, a fully immunocompetent animal model that enables studying viral persistence, protective immunity, and immune-mediated pathogenesis is urgently needed. Since 2011, HCV homologs were identified in a wide range of hosts, including bats, bovines, primates, birds, reptiles, and equids (Hartlage, Cullen, and Kapoor 2016). The genus now includes more than 14 members and is still expanding (International Committee on Taxonomy of Viruses 2021). These newly identified hepaciviruses might serve as useful surrogate animal models and could provide new insights into the evolutionary origins of this virus genus.

The equine hepacivirus (EqHV, Hepacivirus A) is the closest known homolog of HCV (Kapoor et al. 2011). Both viruses share common features including an ∼9.7 kb (+) ssRNA genome which encodes a polyprotein that is cleaved into three structural proteins (Core, E1, and E2) and seven non-structural proteins (p7, NS2, NS3, NSA, NS4B, NS5A, and NS5B) (Kapoor et al. 2011). Amino acid sequence homology between the two viruses is maximal 50–65 per cent within NS3 and NS5B and least within E1E2, NS2, and NS5A (35–45 per cent) (Kapoor et al. 2011). Moreover, both viruses were shown to be hepatotropic (Pfaender et al. 2015, 2017; Ramsay et al. 2015), inducing immune-mediated liver damage (Pfaender et al. 2017; Tegtmeyer et al. 2019) and causing acute as well as chronic infections (Ramsay et al. 2015; Gather et al. 2016). Horses seem to be frequently infected, with a seroprevalence of up to 35 per cent (Burbelo et al. 2012; Pfaender et al. 2015), but the infection is self-limiting in most cases, as indicated by relatively low rates of RNA positivity (Pfaender et al. 2015). Conservation of virological features, including interaction with miR122 (Tanaka et al. 2014; Yu et al. 2017), rapid increase in viremia, and delayed seroconversion were also demonstrated for both viruses (Ramsay et al. 2015; Pfaender et al. 2017).

In this study, we compared E1E2 intra-host diversity in three different cohorts: two previously characterized EqHV-infected equid cohorts (Pfaender et al. 2015, 2017) and an HCV cohort including three patients with either acute or chronic infection without treatment and chronic infection including interferon treatment. For the three cohorts, we characterized longitudinal intra-host evolution and highlighted differences between horses with self-limiting and chronic EqHV infection as well as differences between EqHV and HCV. Additionally, we examined the effect of EqHV transmission and compartmentalization within serum and liver on the virus population structure. Finally, we present evidence that the HVR1 is a unique feature of HCV, that is not conserved in other hepaciviruses.

2. Materials and methods

2.1 Cohort and ethical statement

This study includes serum samples from three longitudinal monitored cohorts: A cohort of five naturally EqHV-infected horses, two of which had a self-limiting infection (horse A—B) and three which were chronically infected (horse C—E). The second cohort includes four naïve horses that were experimentally challenged with EqHV-positive serum drawn from horse C, week 3 (horse F—I). And finally, a cohort of three HCV-infected human patients (Patient J-L). Patient J was acutely infected and cleared infection, patient K was chronically infected and previously treated with IFN, and patient L was chronically infected but rejected treatment. For all individuals, serum samples were drawn regularly (Fig. 1) to measure hepaciviral RNA-levels and for E1E2 deep-sequencing. Furthermore, a liver biopsy and serum sample from an 8-year-old thoroughbred mare were taken to compare virus populations circulating in the serum to resident species in the liver (Pfaender et al. 2015). The serum sample and liver biopsy were taken 7 days apart. Animal experiments were first examined by the animal welfare representatives of the University of Hannover Foundation, and then approved by the Lower Saxony's official authorities (LAVES 13/1262). HCV-positive serum samples were pseudonymously and retrospectively provided by the Department

Downloaded from https://academic.oup.com/ve/article/8/1/veac007/6520266 by Vetmeduni Vienna user on 07 March 2022

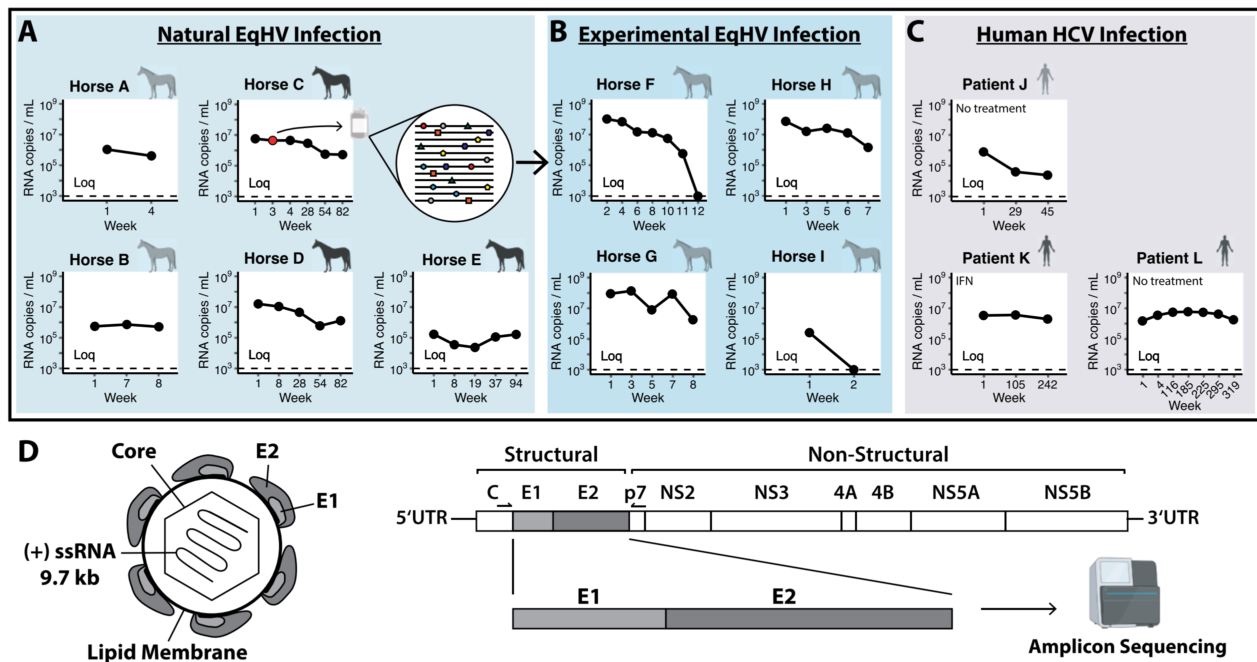

Figure 1. Cohort and study design. (A) For this study, we monitored five naturally infected horses for up to 94 weeks. Two of the horses cleared the infection within 8 weeks (horse A–B) and three horses were persistently infected (horse C–E) as indicated by RNA copy numbers (Loq, limit of quantification). (B) The serum of horse C, week 3 (red dot), was used to inoculate four EqHV-naïve equids which all developed an acute infection (horse F–I). Cohort A and B were previously characterized in (Pfaender et al. 2015, 2017). (C) Furthermore, we analysed the course of HCV infection in one acutely infected patient (patient J) and two chronically HCV-infected patients (K, L). Patient K received interferon and was monitored until sampling week 242. Patient L denied treatment and was monitored for 319 weeks. (D) Schematical representation of the structure of hepaciviruses. The genomic region for the surface glycoproteins E1 and E2 was PCR-amplified, and the amplicons were deep-sequenced with a mean coverage of 5.3×10^4 (±SD 3.1×10^4) (see Supplementary Figure S1). In parts created with www.biorender.com.

of Gastroenterology, Hepatology and Endocrinology; Hannover Medical School in accordance with guidelines of the Declaration of Helsinki. Patients gave their informed consent to use clinical data and blood samples for scientific purposes which were reviewed by the ethics committee at the university centre (ethical vote 2148-2014, Hannover Medical School).

2.2 Hepacivirus RNA extraction and genome quantification

For viral RNA copy number quantification the protocols were previously described (Burbelo et al. 2012; Irving et al. 2014; Pfaender et al. 2015). RNA was isolated from serum samples using the High-Pure Viral RNA Kit (Roche, Mannheim, Germany) according to the manufacturer's manual and subsequently used for cDNA synthesis by using the Prime Script RT Master Mix Kit (Takara) with random hexamer primers. The SYBR Premix Ex Taq II kit (Takara) was used for qRT-PCR to quantify viral RNA copies with primers that were designed to specifically bind to the 5′ untranslated region (Table S1). RNA from horse liver was isolated in 1 mL TRIzol (Life Technology, Schwerte, Germany) from 45 mg tissue.

2.3 E1E2 amplicon generation & next generation sequencing

RNA extracted from the serum samples was reverse transcribed using the Transcriptor High Fidelity cDNA synthesis kit (Roche)—except for horse B at week 7 and 8 where superscript III cDNA synthesis kit (Invitrogen) was used—following manufacturer's instructions with E1E2-specific primers (Table S1). To enable direct comparison of viral population heterogeneity, we adjusted each sample to contain a maximum of 2000 viral cDNA copies. The cDNA was then used as a template for a PCR-amplification using the High Fidelity Platinum® Taq DNA Polymerase (Invitrogen) in a nested PCR reaction with partially degenerated primers (Table S1) specifically adjusted for each individual. Next generation sequencing (NGS) was performed on the Illumina MiSeq platform. The mean coverage was 5.3×10^4 with a standard deviation of 3.1×10^4 (Fig. 1D and Figure S1).

2.4 NGS data processing

Raw paired-end reads were quality checked using FastQC and trimmed via Trimmomatic 0.39. Sequence reads were mapped using Tanoti (https://github.com/vbsreenu/Tanoti) and duplicates were removed with the MarkDuplicates command from Picard tools (Broad Institute 2019). The consensus sequence was extracted using the Sam2Consensus tool (https://github.com/vbsreenu/Sam2Consensus). For each individual, we first generated a consensus sequence of the first sample by mapping against their strain-specific reference sequence (HCV GT 1a: NC_004102, or GT 2b: D10988 and EqHV NZP1: KP325401). This consensus sequence was then used as a reference for subsequent data analysis. The source code is available at https://github.com/CompVirolBochum/HepacivirusIntraHostEvolution.

2.5 Diversity analysis

To analyse intra-host diversity, we compared the virus population of each individual and time point to its respective consensus sequence of the first time point of sampling. For the cohort of experimentally infected horses, we compared the population to the inoculum consensus sequence. We employed a variety of bioinformatic tools which were combined in an in-house pipeline. Briefly, DiversiTools was used to calculate variant frequencies and entropy whilst vNvS was used to calculate $d_N d_S$ values

Downloaded from https://academic.oup.com/ve/article/8/1/veac007/6520266 by Vetmeduni Vienna user on 07 March 2022

and codon tables (https://github.com/josephhughes/DiversiTools, https://github.com/rjorton/vnvs). Haplotypes were constructed using CliqueSNV and subsequently used to construct maximum likelihood trees, using the Tamura-Nei substitution model in MEGA11 (Tamura, Stecher, and Kumar 2021). Statistical robustness of resulting clades was assessed using a bootstrap approach (1000 replicates). The tree in Figure S11 was computed on an amino acid alignment using RaxML. Here the substitution model PROTGAMMAWAG was used, and robustness was assessed with the rapid bootstrap model with 1000 replicates. Graphical representation of the maximum likelihood trees was performed using ggTree and Treeio in R (Wang et al. 2019; Yu 2020). Statistical analysis and visualization were computed with in-house R scripts using the ggplot, tidyverse, cowplot, and ggpubr libraries. The source code is available at https://github.com/CompVirolBochum/HepacivirusIntraHostEvolution.

2.6 Sequence alignments

Multiple sequence alignments were computed using clustalW (Sievers and Higgins 2014), MAFFT (Katoh, Rozewicki, and Yamada 2017) and visualized using ggmsa (Lang and Guangchuang 2021). Entropy calculation and percent identity plots were generated using the Bio2cor (https://cran.r-project.org/web/packages/Bios2cor/index.html) package in R. To evaluate protein properties we used the protein scale function in a sliding window approach (window = 9, edge = 1) from the biopython library (Cock et al. 2009). Transmembrane tendency scales for single amino acids were assessed from the expasy website (Zhao and London 2006). Disorder tendency was calculated using the Spot-Disorder2 webserver (Hanson et al. 2019) and transmembrane prediction was done using TMHMM-server (http://www.cbs.dtu.dk/services/TMHMM/). Glycosylation site prediction was performed via the Glycosite tool available through the LANL database (https://hcv.lanl.gov/content/sequence/GLYCOSITE/glycosite.html) (Zhang et al. 2004).

3. Results

3.1 Intra-host single nucleotide variant diversity in naturally infected horses with acute or chronic EqHV infection

For HCV, virus variability and diversification during the acute phase of infection plays a crucial role in progression to chronicity. We, therefore, investigated evolutionary patterns between horses with a self-limiting infection versus chronically infected horses (Fig. 1A). The intra-host single nucleotide variant (iSNV) frequencies and entropies were calculated for each horse at each time point of sampling over the entire E1E2 sequence in reference to the consensus sequence of the respective first time point (Fig. 2A, B). Cut-off values for high frequency (HF) variants were set at 3.287 per cent (median error [%] plus three times standard deviation of MiSeq Illumina platform (Stoler and Nekrutenko 2021)), while variants above 50 per cent indicate consensus sequence changes. For horses with a self-limiting infection (horse A and B, depicted in blue) iSNV frequencies remained below the HF-threshold and the calculated Shannon entropy showed no elevated peaks across the entire E1E2 sequence (Fig. 2A, B). EqHV in chronically infected horses (horse C—E, depicted in red) showed multiple sites with HF-variants, changes in the consensus sequence at later time points, and the Shannon entropy for these samples exhibited peaks at multiple positions (Fig. 2A, B). Of note, we did not observe any obvious shared mutational hotspots in viral populations isolated from different animals.

We further characterized evolutionary patterns by summarizing diversity per time point and host. Figure 2C shows the total number of iSNVs above 3.287 per cent (black line) as well as the ratio between HF-variants (outer pale dot, set to 1) and consensus changes (inner dot, above 50 per cent). Viral heterogeneity was demonstrably higher in chronically infected horses, with higher numbers of iSNVs, especially in HF-variants, compared to horses with self-limiting infection. These data highlight

Downloaded from https://academic.oup.com/ve/article/8/1/veac007/6520266 by Vetmeduni Vienna user on 07 March 2022

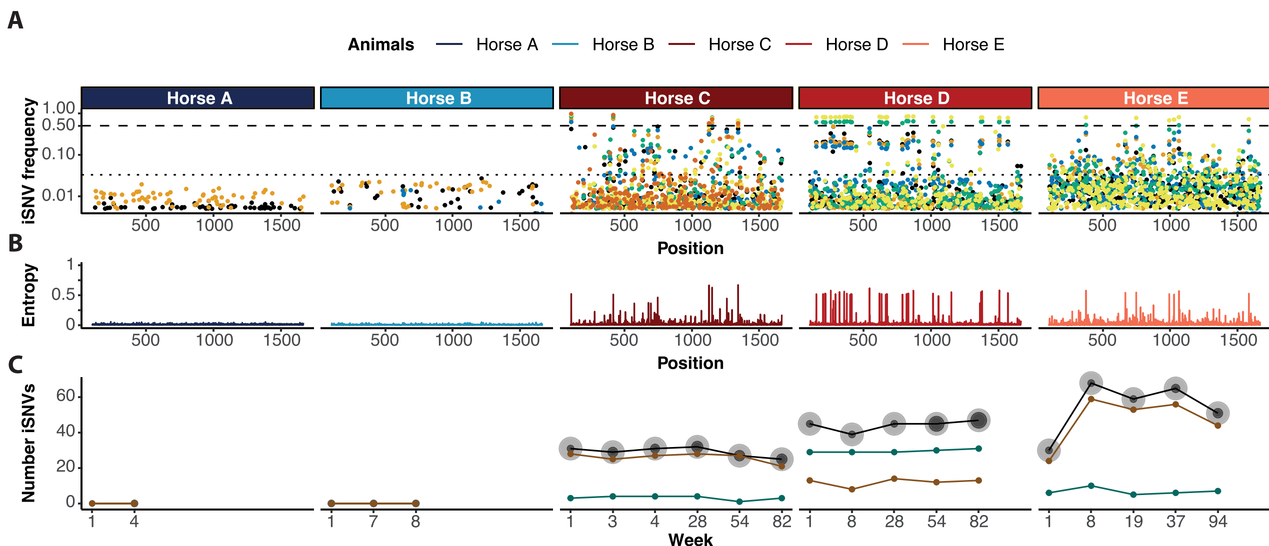

Figure 2. Longitudinal analysis of hepacivirus diversity in naturally infected horses. Acutely infected horses are represented in shades of blue (horse A, B) and chronically infected horses (horse C—E) in shades of red. (A) Variant plots showing the intra-host single nucleotide variants (iSNV) per genomic position and timepoint (represented with dot colour: 1st black, 2nd orange, 3rd blue, 4th green, 5th yellow, 6th red) for naturally infected horses (horse A—E). The dashed lines represent thresholds for HF-variants at a frequency of 0.03287 (3.287 per cent) and 0.5 (50 per cent). (B) Normalized Shannon Entropy per genomic position for each individual over all time points. (C) Number of variants (iSNVs) per time point. HF-variants are depicted in black. The dots represent the ratio of all HF-variants (outer dot) to changes in the consensus sequence (variants above 50 per cent, inner dot). The brown line represents synonymous and the green line non-synonymous substitutions (above 3.287 per cent).

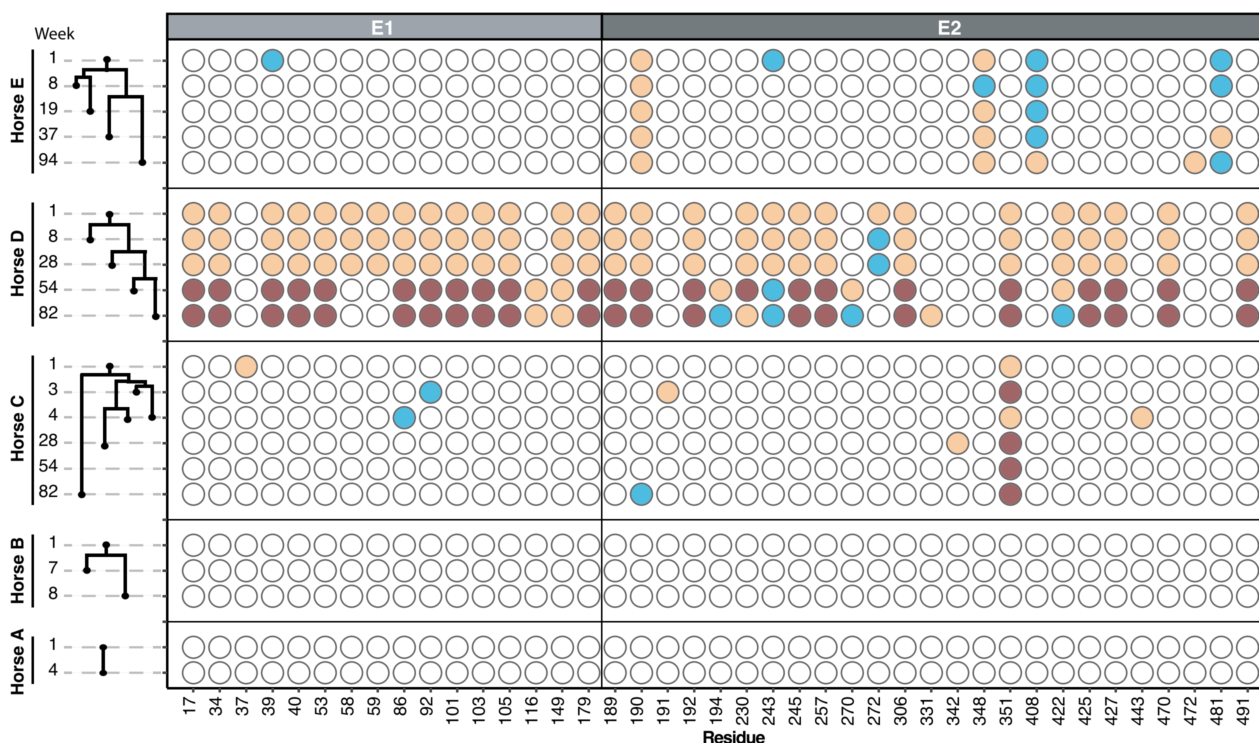

Figure 3. EqHV variant profile in naturally infected horses. Horses are indicated at the y-axis including their sampling time points and phylogenetic clustering. Amino-acid variants above 10 per cent frequency for each position are represented within the dot plot. Sites with a low frequency (below 3.287 per cent) are transparent, those with a medium frequency (between 3.287 per cent–10 per cent) are shown in blue, those with high frequency (up to 50 per cent) in yellow, while changes in the consensus sequence are indicated in red. The bar indicates the position of the variants within E1 and E2.

differences in EqHV mutation accumulation between chronic and acute infection.

Along the total number of variants, we plotted the number of HF-synonymous (brown, above 3. 287 per cent) and HF-non-synonymous (green, above 3. 287 per cent) substitutions to illustrate selection pressure (Fig. 2C). We did not detect any non-synonymous substitutions in the acutely infected horses while in chronically infected horses multiple non-synonymous substitutions were detectable. Interestingly, horse E displayed more non-synonymous than synonymous substitutions. Furthermore, we did not identify pronounced differences between E1 and E2 (Figure S2). Overall, hepacivirus E1E2 diversity was higher in chronically infected than acutely infected horses at the nucleotide level.

Additionally, we compared possible escape mutations from host humoral responses, highlighting substitutions with at least 10 per cent frequency at the amino acid level. These data reveal that E1E2 in chronically EqHV-infected horses was indeed more variable than during an acute infection (Fig. 3). Of note, EqHV in horse D showed an increased number of positively selected sites and more variants at the amino acid level than the EqHV virus population in other horses. Furthermore, a shift in the dominant variant occurred in this animal between weeks 28 and 54 which coincides with decreasing viremia, until week 54. Viral titers then again increased towards the last sampling time point (week 82).

Overall, our data demonstrate virus populations in horses with a self-limiting infection exhibit less E1E2 diversity than viruses isolated from chronically infected horses, indicating that EqHV intra-host diversity may play a role in progression to chronicity.

3.2 Variability and transmission bottleneck in experimentally infected horses

Studying the early events of HCV infection remains a challenging task. Rarely, HCV viremia is detected during the acute phase of infection with knowledge about the inoculum. Although E1E2 tracking studies were performed previously in human chimeric liver mice (Brown et al. 2012), these mice were immune-deficient. In contrast, our study highlights hepacivirus evolution at transmission in a fully immune-competent animal model which supports possible chronic infection (Fig. 1B). Surrogate models offer the opportunity to investigate the events during early infection including transmission bottleneck effects and virus adaptation to a new host. For this, we analysed the genetic composition of virus populations of four EqHV-naïve horses that were inoculated with the serum of horse C (week 3) (Pfaender et al. 2017) and referenced it to the inoculum (Fig. 1B). Horses F and I were infused with 500 mL of one aliquot (inoculum I), horses G and H with 100 mL of a second aliquot (inoculum II), possibly representing diverse snapshots of the viral population. All horses cleared the infection between weeks 7 and 12, except for horse I which was PCR-negative 2 weeks post-inoculation (Fig. 1B). In all experimentally infected horses, EqHV accumulated variants within the new host including changes in the consensus sequence (Fig. 4A). Notably, at each time point, synonymous substitutions were more abundant than non-synonymous and therefore only a few amino acid changes were present (Figs 4 and 5). Again, the number of synonymous and non-synonymous substitutions did not differ markedly between E1 and E2 (Figure S3). Even though the experimentally infected horses developed an acute self-limiting infection, iSNV frequencies, entropy, and the total number of

Downloaded from https://academic.oup.com/ve/article/8/1/veac007/6520266 by Vetmeduni Vienna user on 07 March 2022

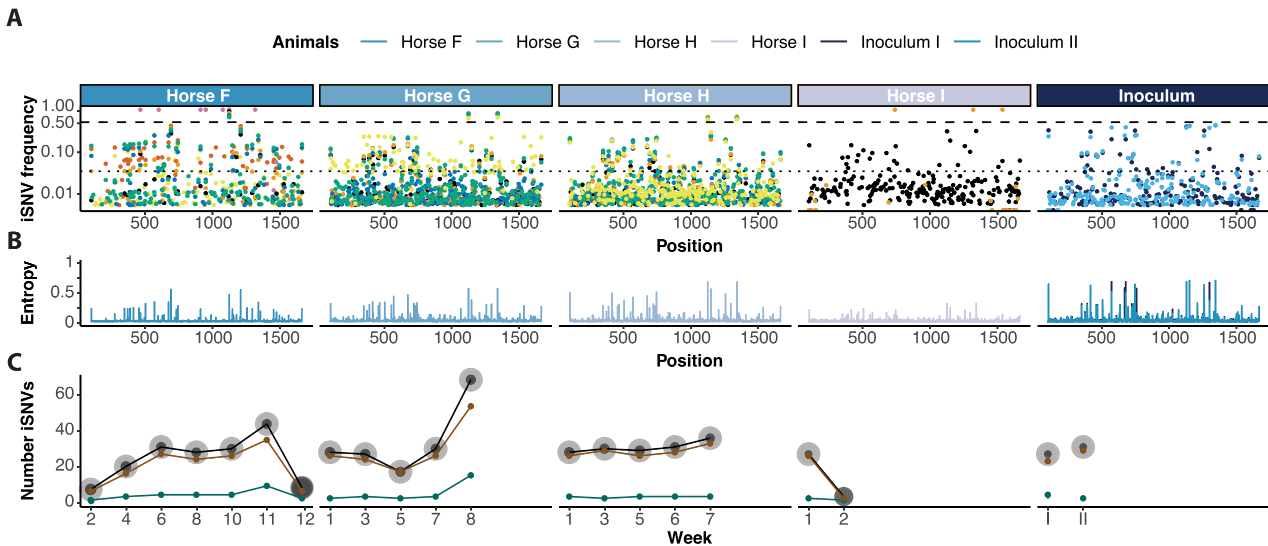

Figure 4. Diversification of EqHV in experimentally infected equids. The serum of horse C was used to inoculate four naïve horses, all of which developed an acute infection (horse F—I). The inoculum was drawn at week 3 and separated into two aliquots which were both deep sequenced (Inoculum I, II). (A) Genetic diversity was characterized by plotting the iSNV frequency for each sample and timepoint (represented with dot colour: 1st black, 2nd orange, 3rd blue, 4th green, 5th yellow, 6th red, 7th pink) to the genomic region of E1E2. Dashed lines represent a 0.03287 (3.287 per cent) and 0.5 (50 per cent) frequency threshold. (B) Normalized Shannon entropy for each sample over all time points. (C) Number of iSNVs per week. HF-variants are depicted in black. The dot size represents all HF-variants (outer dot) and changes in the consensus sequence (variants above 50 per cent, inner dot). The brown line shows synonymous and the green line non-synonymous substitutions (above 3.287 per cent).

iSNVs were higher than in naturally infected horses with a self-limiting infection, which might be linked to the high viral dose and the already high population heterogeneity present in the inoculum (Fig. 4). In general, but especially for horse I, the founder population—here, the first population which was sequenced, demonstrated a high number of iSNVs and iSAVs (intra-host single amino acid variants), indicative of a heterogenous viral population initiating the new infection (Figs 4 and 5). Notably, horse I cleared the infection 2 weeks post-inoculation. In contrast, horse F had the lowest number of iSNVs in the founder population, but throughout the measurement period demonstrated the highest number of consensus sequence changes (Figs 4 and 5) and had the longest viremia (12 weeks). In summary, experimentally infected horses had similar numbers of sites exhibiting positive or negative selection, but a higher number of variants than the virus in naturally (acutely) infected horses.

There are multiple barriers for a virus to establish an infection within a new host, possibly resulting in population bottlenecks. We, therefore, characterized the bottleneck severity and virus population diversification in the experimentally infected horses (Fig. 6). We generated a maximum likelihood tree from reconstructed haplotypes that were generated from periodically drawn serum samples from the experimentally infected horses and the inoculum (Fig. 6 A–D). Notably, haplotype reconstruction is a computational estimation based on nucleotide frequencies and an *in silico* assumption about genome linkage. The inoculum is represented in almost all clades, also clustering with populations from late time points, indicating low bottleneck severity. The overall genetic distance between samples was low, reflecting the close relationship to the inoculum. Subsequently, we compared the number of HF-iSNVs that occur in both the inoculum and the experimentally infected horses to the number of mutations that were unique to the respective experimentally infected horses (Figure S4). While a substantial number of mutations is shared with the inoculum at early time points, over time unique variants arise. However, only a few minor variants impacted the amino acid composition, some of which were already present in the inoculum (Fig. 5). Taken together, there is limited evidence for a transmission bottleneck upon experimental inoculation of horses, and we did not observe a striking reduction in population heterogeneity, which might be linked to the high-dose infection.

Additionally, we evaluated intra-host population bottlenecks by analysing EqHV populations from a liver specimen and a serum sample drawn 7 days apart from an 8-year-old thoroughbred mare (Figure S5). Both specimens were similarly heterogeneous and had HF-variants at identical positions suggesting that both share the same origin. There was no evidence for compartmentalization of E1E2 sequences based on sample origin, suggesting the liver is the only site of active EqHV replication in horses.

3.3 Intra-host diversity of HCV

We studied glycoprotein diversification in a small HCV cohort (Fig. 1C) which included three patients with different treatments and disease statuses. Patient (J) was acutely infected and did not receive antiviral treatment. Patient K and L were both chronically infected. While Patient K was treated with interferon, Patient L rejected treatment. Patient J, therefore, resembled the treatment and disease status of the horses with self-limiting infection and Patient L that of chronically infected horses. HCV variability in acutely infected patient was lower than in chronically infected patients (Fig. 7). Interestingly, this pattern resembles the findings in EqHV-infected horses with a self-limiting infection compared to chronically infected horses. Both chronically infected patients demonstrated high diversity during the sampling period with up to 330 iSNVs and a high number of consensus sequence changes. The number of synonymous substitutions was always higher than the number of non-synonymous substitutions. Moreover, when divided for E1 and E2, the total number of iSNVs was similar, but there were more non-synonymous substitution within E2 than E1 (Figure S6). Furthermore, the number of mutations

Downloaded from https://academic.oup.com/ve/article/8/1/veac007/6520266 by Vetmeduni Vienna user on 07 March 2022

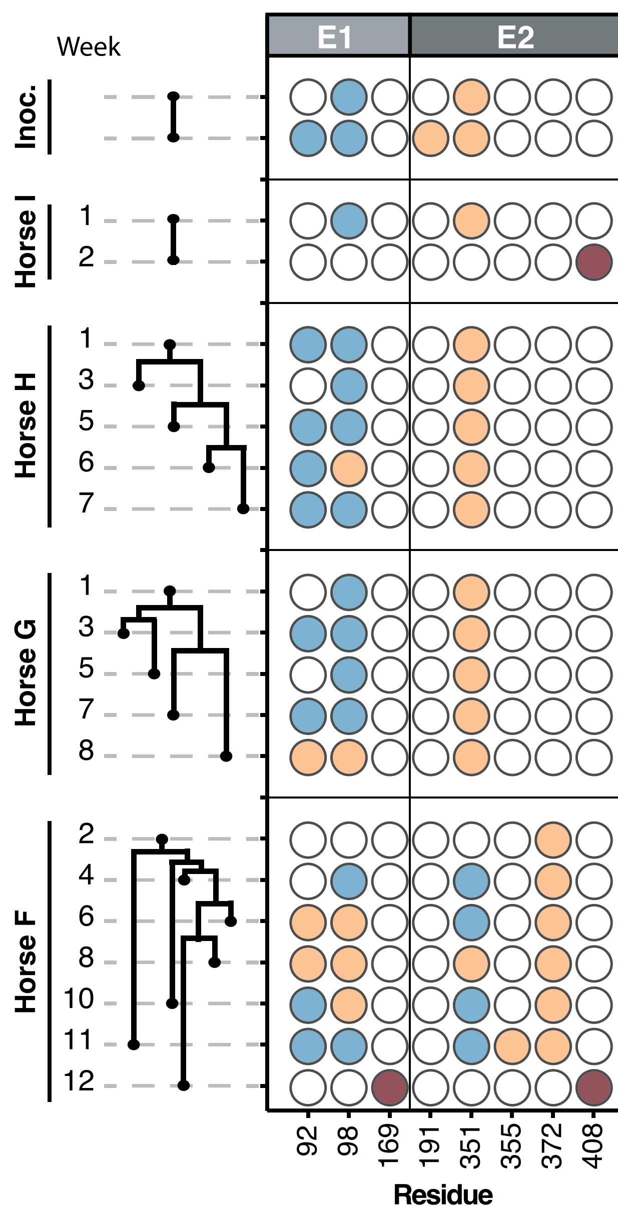

Figure 5. EqHV variant profile in experimentally infected horses. All variants above 10 per cent were isolated and are display in the dot plot. Sites with amino acid frequency below 3.287 per cent are transparent, sites between 3.287 per cent to 10 per cent are shown in blue, high frequency variants are shown in yellow (up to 50 per cent) and consensus sequence changes are depicted in red. Samples are sorted for each animal and time point on the y-axis, including a phylogenetic clustering.

that became fixed in the population increased over time compared to the reference population, and the number of amino acid variants was higher in E2 than in E1 (Fig. 8). Subsequently, we analysed mutational hotspots within EqHV and HCV sequences, by calculating the consensus sequence frequency in a sliding window approach for each animal (Fig. 9A). While the three HCV patients showed high diversity within the N-terminal region of E2, corresponding to the HVR1, EqHV samples did not demonstrate obvious regions with higher variability. This was also apparent when comparing the proportion of amino-acid variants occurring in this region for the different cohorts (Figs 3, 5 and 8). For HCV, 15/62 mutations occurred in E1, 20/62 within the HVR1, and 27/62 within E2, excluding the HVR1. For EqHV in naturally infected horses, 16/41 mutations occurred in E1, 5/62 within the 27 N-terminal amino acids and 20/62 in E2, excluding the 27 N-terminal amino acids. Additionally, we visualized the mutability (longitudinal mutation frequency for all sites potentially mutating) at the amino acid and nucleotide level for E1, E2 and the N-terminal 27 amino acids, which showed that the HVR1 in HCV was a mutational hotspot—this feature is absent in EqHV (Figure S7). Moreover, the differences in variability were more pronounced at the amino acid level than on the nucleotide level, indicating differences in selection pressure between the two viruses in their respective hosts. Supporting this, substitution rates were higher for HCV-infected patients than for EqHV—except horse D (Figure S8).

To conclude, patterns of variability were similar between acutely and chronically infected individuals included in this study in both EqHV and HCV. However, HCV demonstrated higher overall variability than EqHV, especially regarding the N-terminal region of E2. Here, EqHV seems to lack the hypervariability observed within the HVR1 of HCV.

3.4 Hepacivirus glycoprotein evolution

The genomic region for E1E2 differs in size between EqHV and HCV (Figure S9) which is why we constructed an amino acid multiple sequence alignment (MSA) for this region and calculated the percent identity (Fig. 9B). Both aligned well within E1 and the C-terminal E2, while the 34 N-terminal amino acids of E2 did not align to HCV. This region corresponds to the HVR1 (27 amino acids) and the AS412, an epitope targeted by neutralizing antibodies. To evaluate whether this was only a characteristic of the EqHV sequences from our cohort, we retrieved all available EqHV sequences from NCBI where the complete coding region for E1E2 was present (Figure S10). The homology between all sequences was high, demonstrating that our cohort is indeed a representative for this virus. Moreover, entropy values peaked at four positions, including amino acid residue 40 within E1, the N-terminal region of E2 (residue 188) as well as residues around position 255 and 430. Interestingly, the junction between E1E2 showed a length polymorphism of one to four amino acids (Figure S10).

Next, we calculated the disorder tendency for E2 which indicates whether a protein has a fixed, organized three-dimensional structure (disorder tendency < 50) or is rather flexible, ergo unstructured (disorder tendency ≥ 50). The E2 N-terminus of HCV is an intrinsically disordered region, absent in EqHV sequences, while the remaining E2 parts display comparable disorder tendencies (Fig. 9C). In agreement, Kyte–Doolitle hydrophobicity plots displaying transmembrane tendencies for E1 and E2 aligned well outside the N-terminal region of E2 (Fig. 9D).

Since the origin of HCV remains elusive, we compared whether other viruses from the genus *Hepacivirus* had similar features. We computed a MSA and a phylogenetic tree from E1E2 sequences from hepaciviruses that infect mammals (Figure S11 and Supplementary Material). We included the bushbaby hepaci-like virus (BbHV), a recently identified virus which is closely related to HCV and EqHV. Bushbabies (*Galago selegalensis*) are small primates, making them a potential surrogate model. The NCBI-deposited sequence, however, did not include an open reading frame for E1E2 and we therefore re-assembled SRR361358 and manually curated the BbHV genome (see Supplementary Material). The assembled sequence clustered closely to HCV and EqHV (Figure S11). Interestingly, all hepacivirus sequences included in the analysis lacked a robust alignment within the HVR1 of HCV, while E1 as well as the C-terminus of E2 aligned better, implying that the

Downloaded from https://academic.oup.com/ve/article/8/1/veac007/6520266 by Vetmeduni Vienna user on 07 March 2022

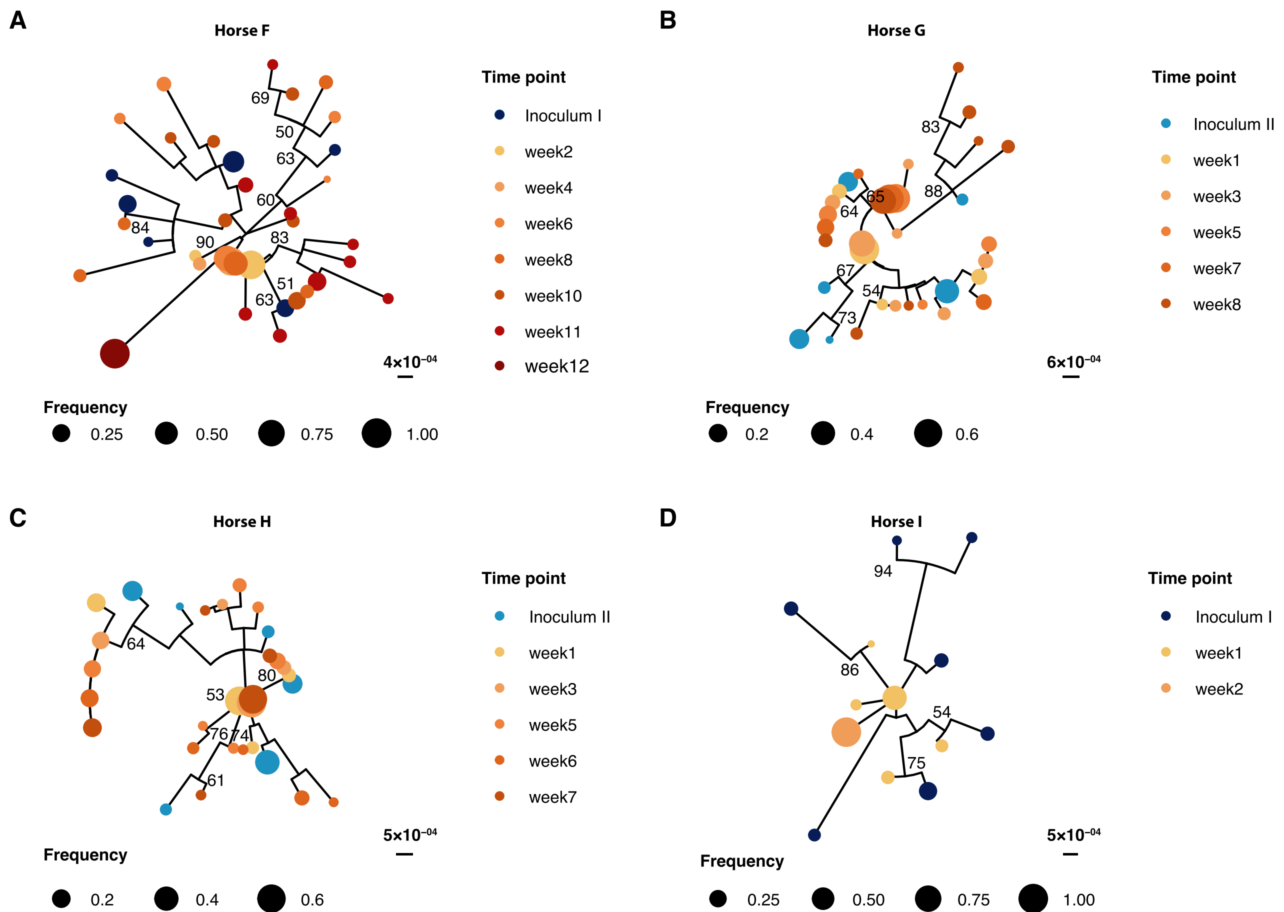

Figure 6. Phylogenetic analysis of hepacivirus diversification in experimentally infected equids. (A–D) Maximum likelihood trees computed from haplotypes including 1000 bootstraps (percent support shown in the tree). The tip colors depict the inoculum (inoculum I, dark blue; inoculum II light blue) and the sample weeks in a yellow to dark red color gradient. The tip size indicates the frequency of each haplotype.

E2 N-terminal region is unique to HCV. Moreover, hepaciviruses infecting rodents (RHV, NrHV) seem to be highly variable, clustering in multiple branches and showing larger misalignments in the MSA (Figures S11 and S12).

In summary, EqHV and HCV showed similar properties outside the E2 N-terminal region which seems to be the case for many hepaciviruses. They typically had a smaller E2 protein than HCV, which, furthermore, did not align to the HVR1 region of HCV.

4. Discussion

HCV remains a global health burden with no robust animal model available to study *in vivo* processes of protective immunity, viral persistence, and pathogenesis. Closely related viruses from the genus *Hepacivirus* could therefore be useful surrogate models to address these questions. Here, we present a whole glycoprotein ultra-deep-sequencing analysis of HCV and EqHV populations from a well-documented longitudinal cohort of nine EqHV-infected horses and three HCV-infected patients, allowing us to characterize similarities and differences in hepacivirus glycoprotein evolution and population composition in the respective hosts.

Intra-host diversity plays a significant role for pathogens to establish infections, evade immune pressure and establish persistency. The first aim of our study was therefore to characterize intra-host diversity in naturally infected horses with either a self-limiting or chronic EqHV infection, using Illumina deep-sequencing. Considering the rather short monitoring period of hepacivirus population dynamics in this study compared to the evolutionary history of these viruses, we decided for the amplification and sequencing of the highly variable glycoproteins E1E2. Prior to amplification, we reverse transcribed the RNA with hepaci-specific primers, which of note, could bear the risk of preferential reverse transcription as well as amplification and may lead to underrepresentation of specific sequences. The number and frequency of EqHV variants were lower in animals with self-limiting infection compared to chronically infected horses, suggesting that variability might play a role in progression to chronicity for EqHV. This is in line with HCV, where sequence variability during the acute phase of infection was described as predictive of transition into chronicity (Enomoto et al. 1993; Sakamoto et al. 1994; Farci et al. 2000), followed by continued sequence diversification during the chronic phase of infection (Ogata et al. 1991; Weiner et al. 1992; Kurosaki et al. 1993; McAllister et al. 1998). There was, however, little evidence for immune evasion mutants within E1 and E2 of chronically infected horses, except for horse D, which showed multiple sites under positive selection pressure. While the pattern of selection pressure in horse D was comparable to the patients in the HCV cohort, the indicators for diversity (number of iSNVs) were much lower. In general, EqHV had a lower intra-host diversity than HCV in the three patients and we did not observe shared mutational hotspots within E1E2.

Downloaded from https://academic.oup.com/ve/article/8/1/veac007/6520266 by Vetmeduni Vienna user on 07 March 2022

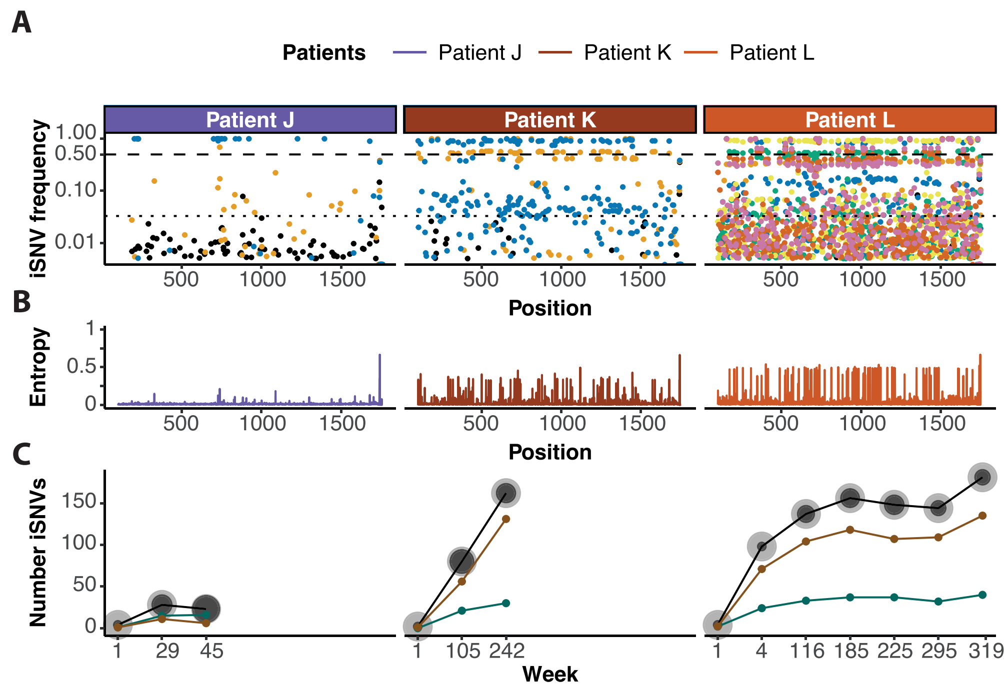

Figure 7. Sequence variability in HCV-infected patients. Sequence variability was compared between an acutely infected patient (J, purple), a chronically infected patient that received interferon treatment (K, red) and a chronically infected patient that rejected treatment (L, orange). (A) Diversification was analysed by plotting iSNVs from each patient and timepoint (represented with dot color: 1st black, 2nd orange, 3rd blue, 4th green, 5th yellow, 6th red, 7th pink) to their genomic position within E1E2. Dashed lines represent a 0.03287 (3.287 per cent) and 0.5 (50 per cent) variant threshold. (B) Normalized Shannon entropy for E1E2 of each individual over all time points. (C) Number of nucleotide variants per week. In black, all HF-variants. The dot size represents all variants above 3.287 per cent (outer dot) and consensus changes (variants above 50 per cent, inner dot). Synonymous and non-synonymous substitutions above 3.287 per cent are shown in brown and green, respectively.

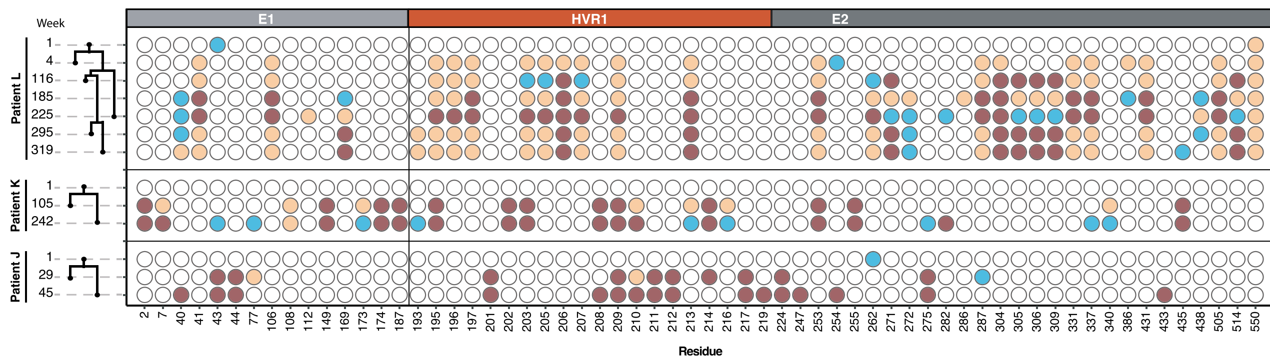

Figure 8. HCV amino acid variants. Residues with variant frequency of at least 10 per cent were isolated and depicted in the dot-plot. Dots in opal represent sites with a mutation frequency below 3.287 per cent, sites in blue are between 3.287 and 10 per cent, sites up to 50 per cent are depicted in yellow and consensus sequence changes are shown in red. The y-axis shows the phylogenetic clustering for each patient sample.

The absence of strong selection pressure and the low diversity could be linked to a weak humoral immune response which has previously been observed in EqHV-infected horses (Pfaender et al. 2017; Tomlinson et al. 2021). Similarly, the absence of immune-mediated pressure in immunocompromised HCV patients and in chimpanzees infected with HCV was found to reduce intra-host glycoprotein diversity (Booth et al. 1998; Bassett et al. 1999; Ray et al. 2000) and, their presence induced a higher rate of non-synonymous substitutions in HCV-infected patients (Farci et al. 2000; Guglietta et al. 2005, 2009). The high diversity in HCV may also be connected to the low-fidelity polymerase NS5B, which has an estimated error rate of 10^{-4} substitutions per site per replication, contributing to intra-host diversity and ultimately the global diversification of HCV into seven genotypes. Whether the EqHV polymerase has a similar fidelity remains to be elucidated, although global patterns of EqHV diversity are substantially lower than seen for HCV, indicating a higher EqHV polymerase fidelity.

Furthermore, E1E2 in HCV has regions under strong selection pressure that therefore exhibit high substitution frequencies, one of them being the HVR1 which is located at the N-terminal region of E2 [reviewed in (Prentoe and Bukh 2018)]. Noteworthy, overall substitution rates were slightly higher than in a previous study, which could be linked to different approaches used for viral sequences detection and calculate evolutionary rates as well as shorter monitoring periods (Raghwani et al. 2019). While high variability was evident within the HVR1 for the HCV cohort in our study, there were no signs of hypervariability within the E1E2 region of EqHV. Since both species are closely related, we

Downloaded from https://academic.oup.com/ve/article/8/1/veac007/6520266 by Vetmeduni Vienna user on 07 March 2022

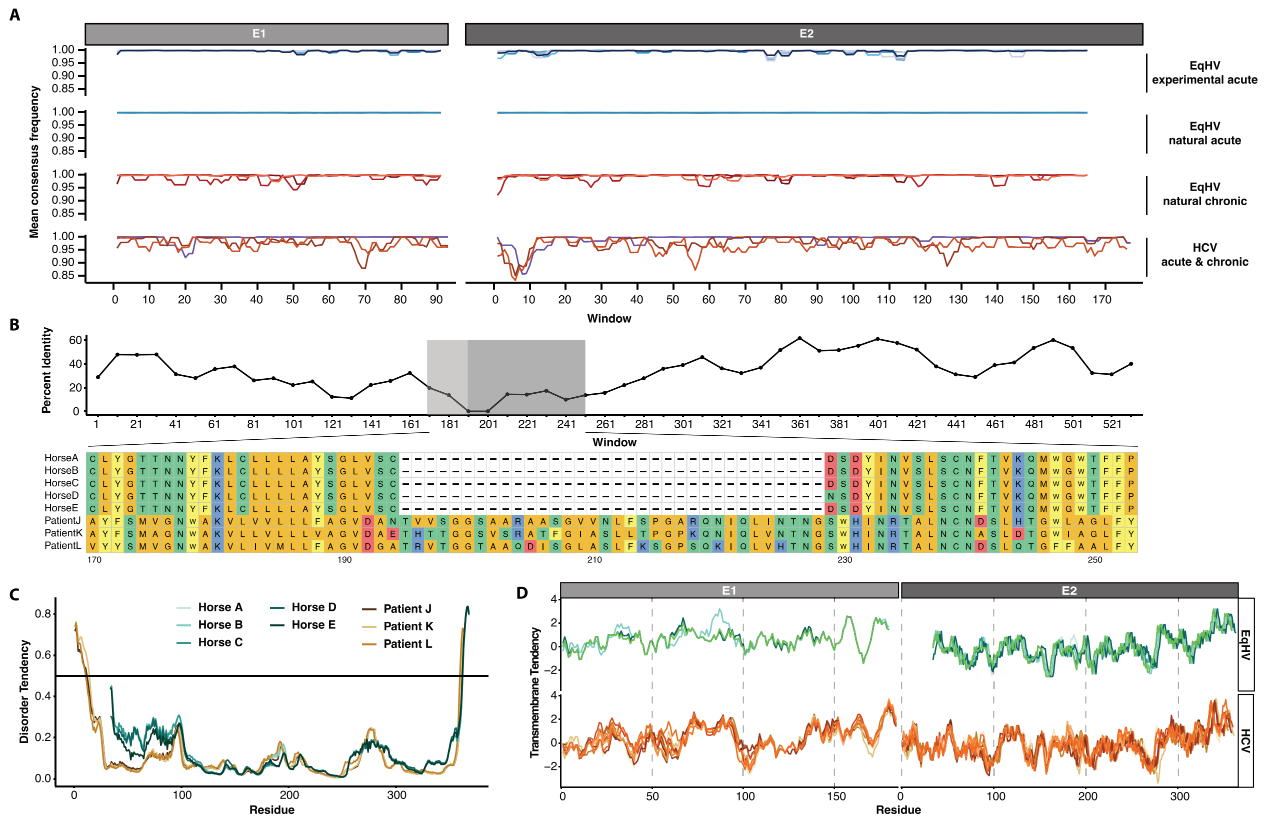

Figure 9. Comparative analysis of the hepacivirus glycoproteins. (A) Sliding window analysis (window size = 20 nt, steps = 6 nt) of the consensus sequence separated for horses with acute infection and horses with chronic infection as well as patients with chronic infection. (B) Conservation analysis of E1E2 between EqHV and HCV. Percent identity was calculated over the E1E2 sequence alignment and the E1-E2 junction was highlighted below, indicating a 34 amino acid misalignment at the N-terminus of E2. (C) Disorder tendency was calculated for E2 using the spot-disorder2 webserver. EqHV sequences were shifted 34 amino acids to the C-terminus. Black horizontal line indicates cut-off for disorder tendency. (D) Kyte–Doolitle hydrophobicity analysis for E1 and E2 for EqHV (green) and HCV (red) sequences. Again, E2 sequnces were shifted 34 amino acids to the C-terminus.

would have expected to find similar genomic features. To investigate this further, we compared the N-terminal regions of E2 from both viruses and observed that the EqHV E2 aligned 34 amino acids into E2 of HCV, roughly reflecting the size difference between E2 in both species (334 amino acids in EqHV and 364 amino acids in HCV). Of note, the size of E1 is relatively similar between the two viruses with only three amino acids difference in length. Furthermore, the previously described disordered tendency of the HVR1 in HCV (Stejskal et al. 2020) seems not to be present in EqHV, while E1 and E2 transmembrane regions were highly conserved. We subsequently compared the sequence homology of hepaciviruses more distantly related to HCV, including viruses that infect bats (Quan et al. 2013), bovines (Baechlein et al. 2015; Corman et al. 2015), rodents (Drexler et al. 2013; Guo et al. 2019) and primates (Canuti et al. 2019), as well as the newly described galago hepacivirus. Amino acid similarity was higher in E1 and the C-terminal region of E2 than in the N-terminal region of E2, which showed larger gaps and misalignment, similar to what we observed in EqHV. Furthermore, for EqHV we noticed a length polymorphism of 1–4 amino acids within the N-terminus of E2, suggesting that this region might be evolutionarily flexible. Hence, we showed that the HVR1 is not present within E2 of EqHV and other hepaciviruses which gives rise to new speculations about the evolutionary origin of HCV and whether HCV acquired this feature *de novo* or whether the other hepaciviruses lost it. The HVR1 seems to be a unique feature for HCV, acting as immunogenic decoy and shielding the conserved CD81 binding site and therefore potentially contributing to its species tropism. To be able to answer questions relating to the functionality of EqHV E2, including the entry mechanism and potential interaction with SRB1, an EqHV cell-culture system is required. The first attempts to generate a cell culture clone were however unsuccessful (Scheel et al. 2015) and replicon or pseudotype-based systems have not been established yet. Additionally, hepaciviruses have narrow host range and tissue tropism and therefore any cell-culture system would need to be based on an equine liver cell-line. Furthermore, it would be interesting to gain insight into how the absence of the HVR1 in EqHV might influence chronicity and the capacity of antibodies to neutralize the virus. Moreover, comparing protein structure between HCV and other hepaciviruses would provide us with a deeper understanding of viral evolution at the protein level. This however remains a difficult task, since the N-terminal region of E2 in HCV is too flexible for experimentally based structure evaluation. Recently, a computationally predicted HCV E2 structure, including the HVR1 but excluding the transmembrane region, was published (Stejskal et al. 2020), which could be useful in overcoming experimental restrictions. Ultimately, the HVR1 seems to be unique feature of HCV and identification of an ancestral HCV-like virus that possesses this region would enable investigation of the evolutionary mechanisms facilitating HCV's successful infection millions of people worldwide.

Downloaded from https://academic.oup.com/ve/article/8/1/veac007/6520266 by Vetmeduni Vienna user on 07 March 2022

Since HCV can only infect humans and chimpanzees, *in vivo* experiments remain a limiting factor for research, especially regarding the early immune responses to HCV infection, as this phase is often asymptomatic and therefore unnoticed. Although a recently published *ex vivo* model was able to highlight early innate immune responses, it still lacks the linkage to adaptive immunity (Tegtmeyer et al. 2021). The surrogate system EqHV in equids allows us to study those early responses to hepacivirus infection and immune-mediated clearance in a controlled environment. In our study, all experimentally infected horses developed a self-limiting acute infection which was cleared between 2 and 7 weeks post-infection. A comparison to the cohort of naturally infected horses is however difficult due to the unknown time point of infection and possible reinfection events. Notably, experimentally infected horses with an acute self-limiting infection demonstrated a higher diversity than horses with a self-limiting infection within the naturally infected cohort. Reinfection with hepaciviruses, including HCV, NrHV (Norway rat hepacivirus) and EqHV were reported to result in improved control of infection, leading to a more rapid and increased clearance (Bassett et al. 2001; Major et al. 2002; Nascimbeni et al. 2003; Lanford et al. 2004; Prince et al. 2005; Osburn et al. 2010; Billerbeck et al. 2017).

Intra-host diversity is important for a virus population to maintain the ability to adapt, especially in the early stages of infection or in response to environmental changes (e.g. seroconversion; Vignuzzi et al. 2006). Usually, during transmission, the virus population loses diversity in a transmission bottleneck, and only a few variants can establish a new infection followed by population diversification, which is known as the founder effect. In contrast, changes in the selective constraints acting upon a viral population at transmission can result in rapid outgrowth of fitter variants in the new host, known as selective sweep. Notably, we did not observe a transmission bottleneck. For HIV and HCV, it was previously shown that transmission bottlenecks occur and that a single founder variant was able to establish new infections, which was followed by population diversification (Keele et al. 2008; Boutwell et al. 2010; Bull et al. 2011). However, in the settings of experimental infection of immune-deficient mice with a patient derived HCV inoculum (Brown et al. 2012), or graft reinfection post liver transplantation, a selective sweep was observed. Interestingly, a study by Bar and colleagues indicated that intravenous drug users were more often infected with multiple HIV founder variants than patients who got infected via sexual intercourse (Bar et al. 2010), which could explain the relatively wide bottleneck post experimental inoculation in our study. Realistically, the large volume used for inoculation (500 mL or 100 mL with 7.78×10^6 virus copies per mL) is likely to be connected to a strong innate immune response—thus despite infection with a diverse population, rapid viral clearance is achieved by induction of strong innate immune responses. Whether low dose infections are sufficient for the experimental inoculation of horses and whether they might even foster persistent infection remains to be investigated. Noteworthy, *ex vivo* infection experiments in primary human hepatocytes indicated that low induction of the innate immune response paved the way for HCV to escape immune surveillance and establish chronicity (Tegtmeyer et al. 2021). Experimentally inoculated adult horses from different studies often develop an acute self-limiting infection (Ramsay et al. 2015; Pfaender et al. 2017; Tomlinson et al. 2021), however, it was shown that foals infected with EqHV remained RNA positive for at least a year post-infection, indicating that persistent EqHV infection can be achieved by experimental inoculation. For experimental infections of rodents with rodent hepacivirus (RHV) it was shown that only ten genome equivalents were sufficient for inducing high viremia while a higher dose of up to 10^4 genome equivalents did not influence virus kinetics (Billerbeck et al. 2017).

After successfully passing the transmission bottleneck and establishing an infection, EqHV showed signs of intra-host diversification. However, increased heterogeneity was mainly observed on nucleotide level, and only a few amino acid substitutions were present. The lack of substitutions could have been connected to rapid clearance during the early phase of infection, between 2 and 12 weeks post-inoculation. This early phase of acute infection plays a critical role in HCV. Early induction of anti-E1E2 antibodies, as well as reduced E1E2 heterogeneity was found to be a determinant of a self-limiting infection (Zibert et al. 1997; Farci et al. 2000; Dowd et al. 2009). Moreover, population bottlenecks can occur at different time points and sites within a host during infection (Joseph et al. 2015). For instance, after seroconversion, immune pressure can drive selection of variants and cause a reduction of diversity or conditions between different compartments can vary, causing some variants to be selected. Here, we were not able to study population bottlenecks in a temporal resolution due to early clearance in the experimentally infected horses. However, we were able to characterize the population structure in both, liver and serum of one chronically infected mare. Compartmentalization in HCV infection seems to be dependent on disease progression and data about the heterogeneity of HCV populations in one host or even one liver is contradictorily discussed (Fan et al. 1999; Sakai et al. 1999; Cabot et al. 2000; Hedegaard et al. 2017). Extra-hepatic reservoir and replication sites for HCV have been identified possibly being explanatory for population differences. Also neutralizing antibodies in the periphery may be indicative for compartmentalization due to overrepresentation of escape variants in patient sera compared to the liver (Cabot et al. 2000). Although so far, EqHV has exclusively been detected in the liver of horses (Pfaender et al. 2015) and our data did not reveal major differences in the viral population composition between the serum and liver specimen, deeper analyses are needed to judge compartmentalization in EqHV infection, considering the availability of data for HCV.

In summary, we showed that sequence variability might be an important feature for chronicity in EqHV, but there was no strong selection pressure detected in infected horses. Furthermore, intra-host variability was less pronounced in EqHV-infected horses compared to HCV-infected patients. Evolutionary analysis revealed sequence features of E1 to remain relatively conserved within the genus of *Hepacivirus*, while E2 was evolutionary more variable, especially with respect to its N-terminal HVR1 which appears to be unique to HCV. EqHV thus represents an interesting model to study mechanisms of virus persistence, evolution and early immune response.

Data availability

Illumina sequencing raw data has been submitted to NCBI repository under BioProject ID PRJNA785969 (http://www.ncbi.nlm.nih.gov/bioproject/785969). The source code is available at https://github.com/CompVirolBochum/HepacivirusIntraHostEvolution.

Supplementary data

Supplementary data is available at *Virus Evolution* online.

Downloaded from https://academic.oup.com/ve/article/8/1/veac007/6520266 by Vetmeduni Vienna user on 07 March 2022

Funding

The study was supported by the Deutsche Forschungsgemeinschaft (German Research Foundation) [grant numbers: 398066876-GRK 2485/1 and 438777365] and the Hungarian Scientific Research Fund [grant number: OTKA/NKFIH FK134311]. T.P. is funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under the Germany's Excellence Strategy – EXC 2155 'RESIST' – Project ID 39087428.

Conflict of interest: None declared.

References

Baechlein, C. et al. (2015) 'Identification of a Novel Hepacivirus in Domestic Cattle from Germany', *Journal of Virology*, 89: 7007–15.

Bankwitz, D. et al. (2017) 'Maturation of Secreted HCV Particles by Incorporation of Secreted ApoE Protects from Antibodies by Enhancing Infectivity', *Journal of Hepatology*, 67: 480–9.

Bar, K. J. et al. (2010) 'Wide Variation in the Multiplicity of HIV-1 Infection among Injection Drug Users', *Journal of Virology*, 84: 6241–7.

Bartenschlager, R. et al. (2018) 'Critical Challenges and Emerging Opportunities in Hepatitis C Virus Research in an Era of Potent Antiviral Therapy: Considerations for Scientists and Funding Agencies', *Virus Research*, 248: 53–62.

Bartenschlager, R., and Lohmann, V. (2000) 'Replication of Hepatitis C Virus', *The Journal of General Virology*, 81: 1631–48.

Bassett, S. E. et al. (1999) 'Viral Persistence, Antibody to E1 and E2, and Hypervariable Region 1 Sequence Stability in Hepatitis C Virus-Inoculated Chimpanzees', *Journal of Virology*, 73: 1118–26.

—— et al. (2001) 'Protective Immune Response to Hepatitis C Virus in Chimpanzees Rechallenged following Clearance of Primary Infection', *Hepatology*, 33: 1479–87.

Billerbeck, E. et al. (2017) 'Mouse Models of Acute and Chronic Hepacivirus Infection', *Science (New York, N.Y.)*, 357: 204–8.

Blum, H. E. (2016) 'History and Global Burden of Viral Hepatitis', *Digestive Diseases*, 34: 293–302.

Booth, J. C. et al. (1998) 'Comparison of the Rate of Sequence Variation in the Hypervariable Region of E2/NS1 Region of Hepatitis C Virus in Normal and Hypogammaglobulinemic Patients', *Hepatology*, 27: 223–7.

Boutwell, C. L. et al. (2010) 'Viral Evolution and Escape during Acute HIV-1 Infection', *The Journal of Infectious Diseases*, 202: S309–S14.

Broad Institute. (2019) 'Picard Toolkit', *Broad Institute, GitHub repository*.

Brown, R. J. P. et al. (2012) 'Hepatitis C Virus Envelope Glycoprotein Fitness Defines Virus Population Composition following Transmission to a New Host', *Journal of Virology*, 86: 11956–66.

—— et al. (2020) 'Liver-expressed *Cd302* and *Cr1l* Limit Hepatitis C Virus Cross-species Transmission to Mice', *Science Advances*, 6: eabd3233.

Bukh, J. (2004) 'A Critical Role for the Chimpanzee Model in the Study of Hepatitis C', *Hepatology*, 39: 1469–75.

Bull, R. A. et al. (2011) 'Sequential Bottlenecks Drive Viral Evolution in Early Acute Hepatitis C Virus Infection', *PLoS Pathogens*, 7: e1002243.

Burbelo, P. D. et al. (2012) 'Serology-enabled Discovery of Genetically Diverse Hepaciviruses in a New Host', *Journal of Virology*, 86: 6171–8.

Burm, R. et al. (2018) 'Animal Models to Study Hepatitis C Virus Infection', *Frontiers in Immunology*, 9: 1032.

Cabot, B. et al. (2000) 'Nucleotide and Amino Acid Complexity of Hepatitis C Virus Quasispecies in Serum and Liver', *Journal of Virology*, 74: 805–11.

Canuti, M. et al. (2019) 'Virus Discovery Reveals Frequent Infection by Diverse Novel Members of the Flaviviridae in Wild Lemurs', *Archives of Virology*, 164: 509–22.

Choo, Q. L. et al. (1989) 'Isolation of a cDNA Clone Derived from a Blood-borne non-A, non-B Viral Hepatitis Genome', *Science*, 244: 359–62.

Cock, P. J. A. et al. (2009) 'Biopython: Freely Available Python Tools for Computational Molecular Biology and Bioinformatics', *Bioinformatics*, 25: 1422–3.

Corman, V. M. et al. (2015) 'Highly Divergent Hepaciviruses from African Cattle', *Journal of Virology*, 89: 5876–82.

Domingo, E., Sheldon, J., and Perales, C. (2012) 'Viral Quasispecies Evolution', *Microbiology and Molecular Biology Reviews*, 76: 159–216.

Dorner, M. et al. (2013) 'Completion of the Entire Hepatitis C Virus Life Cycle in Genetically Humanized Mice', *Nature*, 501: 237–41.

Dowd, K. A. et al. (2009) 'Selection Pressure From Neutralizing Antibodies Drives Sequence Evolution During Acute Infection With Hepatitis C Virus', *Gastroenterology*, 136: 2377–86.

Drexler, J. F. et al. (2013) 'Evidence for Novel Hepaciviruses in Rodents', *PLoS Pathogens*, 9: e1003438.

Enomoto, N. et al. (1993) 'The Hypervariable Region of the HCV Genome Changes Sequentially during the Progression of Acute HCV Infection to Chronic Hepatitis', *Journal of Hepatology*, 17: 415–6.

Fan, X. et al. (1999) 'Comparison of Genetic Heterogeneity of Hepatitis C Viral RNA in Liver Tissue and Serum', *The American Journal of Gastroenterology*, 94: 1347–54.

Farci, P. et al. (2000) 'The Outcome of Acute Hepatitis C Predicted by the Evolution of the Viral Quasispecies', *Science*, 288: 339–44.

Fauvelle, C. et al. (2016) 'Apolipoprotein E Mediates Evasion From Hepatitis C Virus Neutralizing Antibodies', *Gastroenterology*, 150: 206–17.e4.

Gather, T. et al. (2016) 'Acute and Chronic Infections with Nonprimate Hepacivirus in Young Horses', *Veterinary Research*, 47: 97.

Goffard, A., and Dubuisson, J. (2003) 'Glycosylation of Hepatitis C Virus Envelope Proteins', *Biochimie*, 85: 295–301.

Grebely, J. et al. (2012) 'Hepatitis C Virus Clearance, Reinfection, and Persistence, with Insights from Studies of Injecting Drug Users: Towards a Vaccine', *The Lancet Infectious Diseases*, 12: 408–14.

Guglietta, S. et al. (2005) 'Positive Selection of Cytotoxic T Lymphocyte Escape Variants during Acute Hepatitis C Virus Infection', *European Journal of Immunology*, 35: 2627–37.

—— et al. (2009) 'Impact of Viral Selected Mutations on T Cell Mediated Immunity in Chronically Evolving and Self Limiting Acute HCV Infection', *Virology*, 386: 398–406.

Guo, H. et al. (2019) 'Novel Hepacivirus in Asian House Shrew, China', *Science China Life Sciences*, 62: 701–4.

Hanson, J. et al. (2019) 'SPOT-Disorder2: Improved Protein Intrinsic Disorder Prediction by Ensembled Deep Learning', *Genomics, Proteomics & Bioinformatics*, 17: 645–56.

Hartlage, A. S., Cullen, J. M., and Kapoor, A. (2016) 'The Strange, Expanding World of Animal Hepaciviruses', *Annual Review Virology*, 3: 53–75.

Hedegaard, D. L. et al. (2017) 'High Resolution Sequencing of Hepatitis C Virus Reveals Limited Intra-hepatic Compartmentalization in End-stage Liver Disease', *Journal of Hepatology*, 66: 28–38.

International Committee on Taxonomy of Viruses. (2021) *The ICTV Report on Virus Classification and Taxon Nomenclature: Genus Hepacivirus*. <https://talk.ictvonline.org/ictv-reports/ictv_online_

Downloaded from https://academic.oup.com/ve/article/8/1/veac007/6520266 by Vetmeduni Vienna user on 07 March 2022

report/positive-sense-rna-viruses/w/flaviviridae/362/genus-hepacivirus> Accessed 17 Jun 2021.

Irving, W. L. et al. (2014) 'Development of a High-throughput Pyrosequencing Assay for Monitoring Temporal Evolution and Resistance Associated Variant Emergence in the Hepatitis C Virus Protease Coding-region', *Antiviral Research*, 110: 52–9.

Joseph, S. B. et al. (2015) 'Bottlenecks in HIV-1 Transmission: Insights from the Study of Founder Viruses', *Nature Reviews Microbiology*, 13: 414–25.

Kapoor, A. et al. (2011) 'Characterization of a Canine Homolog of Hepatitis C Virus', *Proceedings of the National Academy of Sciences*, 108: 11608.

Katoh, K., Rozewicki, J., and Yamada, K. D. (2017) 'MAFFT Online Service: Multiple Sequence Alignment, Interactive Sequence Choice and Visualization', *Briefings in Bioinformatics*, 20: 1160–6.

Keele, B. F. et al. (2008) 'Identification and Characterization of Transmitted and Early Founder Virus Envelopes in Primary HIV-1 Infection', *Proceedings of the National Academy of Sciences*, 105: 7552–7.

Khera, T. et al. (2021) 'Long-lasting Imprint in the Soluble Inflammatory Milieu despite Early Treatment of Acute Symptomatic Hepatitis C,' *The Journal of Infectious Diseases*.

Kurosaki, M. et al. (1993) 'Rapid Sequence Variation of the Hypervariable Region of Hepatitis C Virus during the Course of Chronic Infection', *Hepatology*, 18: 1293–9.

Lanford, R. E. et al. (2004) 'Cross-genotype Immunity to Hepatitis C Virus', *Journal of Virology*, 78: 1575–81.

Lang, Z., and Guangchuang, Y. (2021), *Ggmsa: Plot Multiple Sequence Alignment Using Ggplot2 R Package Version 1.0.0.* <http://yulab-smu.top/ggmsa/> accessed 1 Dec 2021.

Lavie, M., Hanoulle, X., and Dubuisson, J. (2018) 'Glycan Shielding and Modulation of Hepatitis C Virus Neutralizing Antibodies', *Frontiers in Immunology*, 9: 910–10.

Major, M. E. et al. (2002) 'Previously Infected and Recovered Chimpanzees Exhibit Rapid Responses that Control Hepatitis C Virus Replication upon Rechallenge', *Journal of Virology*, 76: 6586–95.

Manns, M. P., Wedemeyer, H., and Cornberg, M. (2006) 'Treating Viral Hepatitis C: Efficacy, Side Effects, and Complications', *Gut*, 55: 1350–9.

Martell, M. et al. (1992) 'Hepatitis C Virus (HCV) Circulates as a Population of Different but Closely Related Genomes: Quasispecies Nature of HCV Genome Distribution', *Journal of Virology*, 66: 3225–9.

Martinello, M. et al. (2018) 'Management of Acute HCV Infection in the Era of Direct-acting Antiviral Therapy', *Nature Reviews Gastroenterology & Hepatology*, 15: 412–24.

McAllister, J. et al. (1998) 'Long-term Evolution of the Hypervariable Region of Hepatitis C Virus in a Common-source-infected Cohort', *Journal of Virology*, 72: 4893–905.

Meunier, J.-C. et al. (2005) 'Evidence for Cross-genotype Neutralization of Hepatitis C Virus Pseudo-particles and Enhancement of Infectivity by Apolipoprotein C1', *Proceedings of the National Academy of Sciences of the United States of America*, 102: 4560–5.

Nascimbeni, M. et al. (2003) 'Kinetics of $CD4^+$ and $CD8^+$ Memory T-Cell Responses during Hepatitis C Virus Rechallenge of Previously Recovered Chimpanzees', *Journal of Virology*, 77: 4781–93.

Ogata, N. et al. (1991) 'Nucleotide Sequence and Mutation Rate of the H Strain of Hepatitis C Virus', *Proceedings of the National Academy of Sciences of the United States of America*, 88: 3392–6.

Osburn, W. O. et al. (2010) 'Spontaneous Control of Primary Hepatitis C Virus Infection and Immunity against Persistent Reinfection', *Gastroenterology*, 138: 315–24.

Pfaender, S. et al. (2015) 'Clinical Course of Infection and Viral Tissue Tropism of Hepatitis C Virus-like Nonprimate Hepaciviruses in Horses', *Hepatology*, 61: 447–59.

—— et al. (2017) 'Immune Protection against Reinfection with Nonprimate Hepacivirus', *Proceedings of the National Academy of Sciences*, 114: E2430–9.

Prentoe, J., and Bukh, J. (2018) 'Hypervariable Region 1 in Envelope Protein 2 of Hepatitis C Virus: A Linchpin in Neutralizing Antibody Evasion and Viral Entry', *Frontiers in Immunology*, 9: 2146.

Prince, A. M. et al. (2005) 'Protection against Chronic Hepatitis C Virus Infection after Rechallenge with Homologous, but Not Heterologous, Genotypes in a Chimpanzee Model', *The Journal of Infectious Diseases*, 192: 1701–9.

Quan, P. L. et al. (2013) 'Bats are a Major Natural Reservoir for Hepaciviruses and Pegiviruses', *Proceedings of the National Academy of Sciences of the United States of America*, 110: 8194–9.

Raghwani, J. et al. (2019) 'High-Resolution Evolutionary Analysis of Within-Host Hepatitis C Virus Infection', *The Journal of Infectious Diseases*, 219: 1722–9.

Ramsay, J. D. et al. (2015) 'Experimental Transmission of Equine Hepacivirus in Horses as a Model for Hepatitis C Virus', *Hepatology*, 61: 1533–46.

Ray, S. C. et al. (2000) 'Hypervariable Region 1 Sequence Stability during Hepatitis C Virus Replication in Chimpanzees', *Journal of Virology*, 74: 3058–66.

Reig, M. et al. (2016) 'Unexpected High Rate of Early Tumor Recurrence in Patients with HCV-related HCC Undergoing Interferon-free Therapy', *Journal of Hepatology*, 65: 719–26.

Renzulli, M. et al. (2018) 'Imaging Features of Microvascular Invasion in Hepatocellular Carcinoma Developed after Direct-acting Antiviral Therapy in HCV-related Cirrhosis', *European Radiology*, 28: 506–13.

Sakai, A. et al. (1999) 'Quasispecies of Hepatitis C Virus in Serum and in Three Different Parts of the Liver of Patients with Chronic Hepatitis', *Hepatology*, 30: 556–61.

Sakamoto, N. et al. (1994) 'Sequential Change of the Hypervariable Region of the Hepatitis C Virus Genome in Acute Infection', *Journal of Medical Virology*, 42: 103–8.

Scheel, T. K. et al. (2015) 'Characterization of Nonprimate Hepacivirus and Construction of a Functional Molecular Clone', *Proceedings of the National Academy of Sciences of the United States of America*, 112: 2192–7.

Sievers, F., and Higgins, D. G. (2014) 'Clustal Omega, Accurate Alignment of Very Large Numbers of Sequences', *Methods in Molecular Biology (Clifton, N.J.)*, 1079: 105–16.

Simmonds, P. et al. (2017) 'ICTV Virus Taxonomy Profile: Flaviviridae', *Journal of General Virology*, 98: 2–3.

Stejskal, L. et al. (2020) 'Flexibility and Intrinsic Disorder are Conserved Features of Hepatitis C Virus E2 Glycoprotein', *PLoS Computational Biology*, 16: e1007710.

Stoler, N., and Nekrutenko, A. (2021) 'Sequencing Error Profiles of Illumina Sequencing Instruments', *NAR Genomics and Bioinformatics*, 3: lqab019.

Tamura, K., Stecher, G., and Kumar, S. (2021) 'MEGA11: Molecular Evolutionary Genetics Analysis Version 11', *Molecular Biology and Evolution*, 38: 3022–7.

Tanaka, T. et al. (2014) 'Hallmarks of Hepatitis C Virus in Equine Hepacivirus', *Journal of Virology*, 88: 13352–66.

Tegtmeyer, B. et al. (2019) 'Chronic Equine Hepacivirus Infection in an Adult Gelding with Severe Hepatopathy', *Veterinary Medicine and Science*, 5: 372–8.

Downloaded from https://academic.oup.com/ve/article/8/1/veac007/6520266 by Vetmeduni Vienna user on 07 March 2022

—— et al. (2021) ‘Initial Hepatitis C Virus Infection of Adult Hepatocytes Triggers a Temporally Structured Transcriptional Program Containing Diverse Pro- and Antiviral Elements’, *Journal of Virology*, 95: e00245–21.
Tomlinson, J. E. et al. (2021) ‘Pathogenesis, miR-122 Gene-regulation, and Protective Immune Responses after Acute Equine Hepacivirus Infection’, *Hepatology*, 74: 1148–63.
Vignuzzi, M. et al. (2006) ‘Quasispecies Diversity Determines Pathogenesis through Cooperative Interactions in a Viral Population’, *Nature*, 439: 344–8.
von Hahn, T. et al. (2007) ‘Hepatitis C Virus Continuously Escapes from Neutralizing Antibody and T-cell Responses during Chronic Infection in Vivo’, *Gastroenterology*, 132: 667–78.
Wang, L.-G. et al. (2019) ‘Treeio: An R Package for Phylogenetic Tree Input and Output with Richly Annotated and Associated Data’, *Molecular Biology and Evolution*, 37: 599–603.
Weiner, A. J. et al. (1992) ‘Evidence for Immune Selection of Hepatitis C Virus (HCV) Putative Envelope Glycoprotein Variants: Potential Role in Chronic HCV Infections’, *Proceedings of the National Academy of Sciences of the United States of America*, 89: 3468–72.
World Health Organization. (2017) ‘Global Hepatitis Report, 2017’.
Yu, G. et al. (2017) ‘miRNA Independent Hepacivirus Variants Suggest a Strong Evolutionary Pressure to Maintain miR-122 Dependence’, *PLoS Pathogens*, 13: e1006694.
—— (2020) ‘Using Ggtree to Visualize Data on Tree-Like Structures’, *Current Protocols in Bioinformatics*, 69: e96.
Zhang, M. et al. (2004) ‘Tracking Global Patterns of N-linked Glycosylation Site Variation in Highly Variable Viral Glycoproteins: HIV, SIV, and HCV Envelopes and Influenza Hemagglutinin’, *Glycobiology*, 14: 1229–46.
Zhao, G., and London, E. (2006) ‘An Amino Acid “transmembrane tendency” Scale that Approaches the Theoretical Limit to Accuracy for Prediction of Transmembrane Helices: Relationship to Biological Hydrophobicity’, *Protein Science : A Publication of the Protein Society*, 15: 1987–2001.
Zibert, A. et al. (1997) ‘Early Antibody Response against Hypervariable Region 1 Is Associated with Acute Self-limiting Infections of Hepatitis C Virus’, *Hepatology*, 25: 1245–9.

Downloaded from https://academic.oup.com/ve/article/8/1/veac007/6520266 by Vetmeduni Vienna user on 07 March 2022