UAB
Universitat Autònoma de Barcelona

Quasispecies Fitness Partition Analysis of Variability and Resistance Mutations as Predictors of Ribavirin Failure and Persistence in Hepatitis E Virus Infection

Sergi Colomer Castell

Quasispecies Fitness Partition Analysis of Variability and Resistance Mutations as Predictors of Ribavirin Failure and Persistence in Hepatitis E Virus Infection

Sergi Colomer Castell

Thesis submitted for the Degree of Doctor in Biochemistry, Molecular Biology and Biomedicine

Laboratory of Viral Hepatitis

Unit of Liver Diseases

Vall d'Hebron Research Institute

Department of Biochemistry and Molecular Biology

Faculty of Biosciences

Autonomous University of Barcelona

Barcelona, 2026

Thesis Director

Josep Quer

Thesis Tutor

Dr. Jaume Farrés Vicén

Thesis Candidate

Sergi Colomer Castell

The compendium of articles included in this thesis has been approved by the Doctoral Program Academic Commission (CAPD) on 23th of January 2026 following UAB academic regulations and in accordance the RD 1393/2007 of 29 of October, modified by the RD861/2010. These includes:

1. Quasispecies Fitness Partition to Characterize the Molecular Status of a Viral Population. Negative Effect of Early Ribavirin Discontinuation in a Chronically Infected HEV Patient. Josep Gregori†, Sergi Colomer-Castell†, Carolina Campos, Marta Ibañez-Lligoña, Damir Garcia-Cehic, Ariadna Rando-Segura, Caroline Melanie Adombie, Rosa Pintó, Susanna Guix, Albert Bosch, Esteban Domingo, Isabel Gallego, Celia Perales, Maria Francesca Cortese, David Tabernero, Maria Buti, Mar Riveiro-Barciela, Juan Ignacio Esteban, Francisco Rodriguez-Frias and Josep Quer*. Int J Mol Sci 23, 14654 (2022). https://doi.org/10.3390/ijms232314654.

2. In-Host HEV Quasispecies Evolution Shows the Limits of Mutagenic Antiviral Treatments. Sergi Colomer-Castell†, Josep Gregori†, Damir Garcia-Cehic, Mar Riveiro-Barciela, Maria Buti, Ariadna Rando-Segura, Judit Vico-Romero, Carolina Campos, Marta Ibañez-Lligoña, Caroline Melanie Adombi, Maria Francesca Cortese, David Tabernero, Juan Ignacio Esteban, Francisco Rodriguez-Frias and Josep Quer*. Int J Mol Sci 24, 17185 (2023). https://doi.org/10.3390/ijms242417185.

This thesis and the articles derived have been financial supported by Grants PI22/00258 funded by Instituto de Salud Carlos III (ISCIII) and cofounded by the European Union; Centro para el Desarrollo Tecnológico Industrial (CDTI) from the Spanish Ministry of Economy and Business, grant number IDI-20200297 Consorcio Centro de Investigación Biomédica en Red (CIBEREHD).

Sergi Colomer Castell has received financial support from the Ministerio de Ciencia, Innovación y Universidades grant FPU21/04150.

ACKNOWLEDGEMENTS

Finalment, el fruit del treball de tots estos anys ja ha arribat i està ací. Primer de tot, agrair-te Josep el haver estat sempre acompanyant-me durant la tesi. Gràcies a les teues idees, per tots els projectes en els que m'has permès col·laborar. Ja des del màster dèiem que eres un papà científic, i crec que està més que demostrat. Moltes gràcies també al Jaume Farrés, per haver tutelat aquesta tesi i haver-me ajudat sempre que ho he necessitat.

A la Carolina, per haver estat al meu costat des del principi d'entrar en el laboratori. Per totes les bogeries, *rises* i rallades que hem passat junts. A la Marta, vas arribar un poc després però vas ser necessària per completar les *supernenes* afegint el teu to de alegria i *flirty*. Heu sigut les millors companyes de laboratori que podria haver tingut. Si els dies en el laboratori són bons és precisament gràcies a vosaltres dues. Hem viscut mil experiències junts durant tots aquests anys, i que siguen moltíssimes més. I molt de *gossip* més.

Agrair al Damir, "l'americà" per haver-me ensenyat i ajudat a perfeccionar la feina del laboratori. Has sigut un gran mestre. També al Josep Gregori, per la seua gran feina i ajuda amb els anàlisis bioinformàtics, gran part d'aquesta tesis és gràcies a tu.

Moltes gràcies a la Francesca i al Paco, per totes les converses i pluges d'idees. Espere mantindre la passió amb la que parlem de ciència tots junts. També agrair a la Dra. Riveiro, la Dra. Buti, Dr. Ruiz i la Judit per tota la gestió de les mostres, que no han sigut poques. També a tots els participants dels articles, ja que gràcies a la feina de tots aquesta tesi ha eixit endavant.

A tota la gent que ha passat per el laboratori durant este temps, Melanie, Ares, Alvaro, Jorgina, Süm. Hem passat molts bons moments i el laboratori sempre ha sigut un bon punt de trobada. En especial donar les gràcies a la Meritxell per haver-me ensenyat les tècniques amb els gradients i la citometria, que mira que costaven!

Moltes gràcies Zaida per tots els cafenets en els que per un moment podíem desconnectar de la vida. Jo crec que estem destinats a treballar junts, no sé per què serà. Tampoc hauria sigut això possible sense els meus amics de la carrera i màster, Cris, Mauro, David, Laura, Morgane, Raquel, Edu, Jorge, Constan, Lola, Marina.

Gràcies a tots els meus amics que m'han recolzat tant durant estos anys. Als de Tailàndia (a la que crec que mai anirem junts): Lei, Maik, Tama, Carla, Nico, Maite, Paco, Shei, Clarla, Carles, Antoni, Gon, Yoli. Als del grup del cacauet-Barrakueli: Borja, Óscar, Laura, Alba, Marina, Elena. Cada vegada que ens reunim és com si no haguera passat el temps. Joan i Vicent, sempre heu estat quan ho he necessitat i això ho valoraré sempre. A Irene i Joan Andreu per ser els nostres compis de Liceu i de torrà els diumenges. També als meus companys de la xaranga Askampal Poll per tots els bons moments passats i als bandarres de la Valenciana per haver fet estos anys tan divertits.

Moltes gràcies als meus pares Encarna i Jose, i a la meua germana Lorena, per haver-me aconsellat i recolzat amb totes les decisions que he pres durant aquests anys. Dona una tranquil·litat increïble saber que passe el que passe sempre estareu. També a tota la meua família, tios i ties, cosins i nebodet. És una alegria sempre que torne a Alzira. També agrair a Anna i Andreu per tota la ajuda i companyia que m'haveu brindat, que no ha sigut poca.

No podia acabar sense donar les gràcies a Andrea, la meua companya de vida. Soles tu i jo sabem les dificultats que hem pogut passar i com ens hem ajudat incondicionalment. Moltes gràcies per ser com eres i complementar-me tan bé. Saps que t'admire molt i sóc molt afortunat.

TABLE OF CONTENTS

FIGURE INDEX

ABBREVIATIONS

A

AF	Allele Frequency
ALF	Acute Liver Failure
ALT	Alanine Aminotransferase
ASGPR	Asialoglycoprotein Receptors
ATP5B	ATP Synthase F1 Subunit Beta

C

Ctl	Control
CV	Coefficient of Variation

D

DAAs	Direct-Acting Antivirals
DNA	Deoxyribonucleic Acid

E

EASL	European Association for the Study of the Liver
EGFR	Epidermal Growth Factor Receptor
eHEV	Quasi-Enveloped Hepatitis E Virus
EOT	End of Treatment
ESCRT	Endosomal Sorting Complex Required for Transport
ET-NANBH	Enterically Transmitted non-A, non-B hepatitis

F

FHF	Fulminant Hepatis Failure
FLASH	Fast Length Adjustment of SHort reads
FMDV	Foot-and-Mouth Disease Virus
FPV	Favipiravir

G

GRP78	78 kDa Glucose-Regulated Protein
GTP	Guanosine Triphosphate

H

HAV	Hepatitis A Virus
HBV	Hepatitis B Virus
HCV	Hepatitis C Virus
HEV	Hepatitis E Virus
HEV1	Hepatitis E Virus Genotype 1
HEV2	Hepatitis E Virus Genotype 2
HEV3	Hepatitis E Virus Genotype 3
HEV4	Hepatitis E Virus Genotype 4
HEV5	Hepatitis E Virus Genotype 5
HEV6	Hepatitis E Virus Genotype 6
HEV7	Hepatitis E Virus Genotype 7
HEV8	Hepatitis E Virus Genotype 8
HIV	Human Immunodeficiency Virus
HNP	Hill Numbers Profile
Hs	Normalised Shannon Entropy
HVR	Hypervariable Region

I

IMPDH	Inosine 5′-Monophosphate Dehydrogenase
IQR	Interquartile Range
IU	International Units

M

MDS	Multidimensional Scale Plot
Mfmax	Maximum Mutation Frequency
Mfmin	Minimum Mutation Frequency
mNGS	Shotgun Metagenomics Next Generation Sequencing

N

neHEV	Non-enveloped Hepatitis E virus
NGS	Next Generation Sequencing
NS5B	Nonstructural protein 5B

O

ORF	Open Reading Frame
ORF1	Open Reading Frame 1
ORF2	Open Reading Frame 2
ORF3	Open Reading Frame 3

P

PCP	Papain-like cystein protease
PCR	Polymerase Chain Reaction
Pct0.1	Very low fitness QFF. Equivalent to RHL 0.1
Pct1	Low fitness QFF. Equivalent to RHL1
pORF1	Protein ORF1
pORF2	Protein ORF2
pORF3	Protein ORF3

PRP	Proline Rich Region

Q

qD	Hill Numbers
QFF	Quasispecies Fitness Fractions

R

Rab	Ras Associated Binding
RAS	Resistance-Associaed Substitutions
RBV	Ribavirin
RdRp	RNA dependent RNA polymerase
RHL	Rare Haplotype Load
RHL0.1	Very low fitness QFF. Equivalent to Pct0.1
RHL1	Low fitness QFF. Equivalent to Pct1
RNA	Ribonucleic Acid
RSV	Respiratory Syncytial Virus
RTP	Ribavirin Triphosphate
RT-PCR	Reverse Transcription Polymerase Chain Reaction

S

SMRT	Single-Molecule Real-time
SOF	Sofosbuvir
SVR	Sustained Virological Response

T

Th	T Helper
TIM-1	T-cell Immunoglobulin and Mucin Domain 1
Ts/Tv	Transition to Transversion Ratio
Tsg101	Tumor Susceptibility Gene 101

U

UPGMA	Unweighted Pair Group Method with Arithmetic Mean

V

VL	Viral Load

W

WHO	World Health Organisation
WT	Wild Type

SUMMARY

Hepatitis E is an inflammatory liver disease caused by the Hepatitis E Virus (HEV), a leading cause of acute viral hepatitis worldwide with an estimated 20 million annual infections. Although most infections are self-limiting, the disease can progress to a chronic state in immunocompromised individuals, such as solid organ transplant recipients, potentially leading to progressive liver fibrosis and cirrhosis. Current clinical guidelines recommend a three-month course of ribavirin (RBV) monotherapy. However, primary treatment failure and viral relapse remain significant obstacles, particularly due to the lack of approved alternative antivirals.

This thesis investigates HEV as a quasispecies, a dynamic population of closely related but genetically distinct variants that exists as a “mutant cloud”, mainly due to the lack of proofreading mechanisms in the viral polymerase RNA-dependent RNA polymerase. To better understand and visualize these dynamics, we have established a novel analytical approach called Quasispecies Fitness Fractions (QFF), stratifying the viral population into four categories based on haplotype frequency: master, emerging, low fitness and very low fitness. This method provides a more comprehensive visualization of the quasispecies status and evolution rather than traditional diversity metrics, which simplify complex viral structures into single numerical values.

Through the QFF categorization to two HEV long-lasting chronically infected patients who failed multiple courses of RBV, we have identified a significant discrepancy between genotypic and phenotypic evolution when the virus was subjected to RBV pressure. Longitudinal analysis revealed that, while viral populations became highly unstructured at the nucleotide level (genotypic), they remained remarkably stable at the amino acid level (phenotypic), with the master phenotype accounting for 60-80% of the viral population, whereas the master genotype dropped below 10%. These findings suggest that HEV

may take profit of synonymous mutations to preserve protein functionality and viral fitness, effectively using the mutagenic drug as an evolutionary catalyst rather than undergoing lethal mutagenesis or entering to “error catastrophe” characterized by loss of viral genetic information. This increased genetic heterogeneity enables the virus to adapt more rapidly to selective pressures.

To complement quasispecies studies, we applied shotgun metagenomics sequencing to the samples from the patient of the first study who exhibited a profile of RBV resistance development upon a second exposure to the drug. This approach identified several resistance-associated substitutions within the RNA-dependent RNA polymerase (RdRp) amino acid domain located in ORF1 genomic region, including D1384N and Y1587F. Importantly, some of these mutations were detectable as minor variants at sub-consensus levels before they became dominant in the population or contributed to clinical treatment failure. The study also confirmed the increase in diversity observed in the QFF analysis, as well as the rise in the transition-to-transversion ratio (Ts/Tv), serving as a genomic signature of RBV’s mutagenic impact.

From a clinical perspective, our results suggest that RBV monotherapy, particularly in patients with high HEV viral loads, may drive the selection of fitness-enhancing mutations. Therefore, mutagenic therapies should ideally be combined with other antiviral inhibitors able to reduce the size of the mutant cloud and minimize the likelihood of resistance emergence. In addition, to prevent relapse, RBV should be continued for several months until serum HEV-RNA becomes undetectable tests using the most sensitive assay available, if possible.

Finally, this thesis reinforces the clinical importance of deep variant profiling using Next-Generation Sequencing (NGS) as an early predictive tool. By monitoring minor variants at sub-consensus levels, clinicians can anticipate treatment failure in advance, allowing for the early discontinuation of ineffective therapies and the implementation of precision medicine strategies.

The integration of the QFF method and whole genome analysis provides a robust framework for understanding viral persistence and improving therapeutic outcomes in chronic HEV infection.

L'Hepatitis E és una malaltia inflamatòria del fetge causada pel virus de l'Hepatitis E (VHE), una de les principals causes d'hepatitis vírica aguda en el món, amb una estimació de 20 milions d'infeccions anuals. Encara que la majoria de les infeccions són autolimitades, la malaltia pot evolucionar cap a un estat crònic en pacients immunodeprimits, com els receptors de transplantament d'òrgans sòlids, i pot conduir a una fibrosi hepàtica progressiva i a cirrosi. Les guies clíniques actuals recomanen un tractament de tres mesos amb ribavirina (RBV) en monoteràpia. No obstant això, el fracàs terapèutic primari i la recaiguda viral continuen sent problemes importants, sobretot a causa de la manca d'alternatives antivirals aprovades.

En aquesta tesi s'investiga el VHE com un virus estructurat en quasiespècies, és a dir, una població dinàmica de variants estretament relacionades però genèticament diferents que existeixen com un "núvol de mutants", principalment a causa de l'absència de mecanismes de lectura i correcció d'errors en l'ARN polimerasa ARN dependent del virus (RdRp). Per entendre i visualitzar millor aquestes dinàmiques, hem establert un nou enfocament analític anomenat *Quasispecies Fitness Fractions* (QFF), que estratifica la població viral en quatre categories segons la freqüència dels haplotips: *master*, emergent, minoritaris i molt minoritaris. Aquest mètode proporciona una visualització més completa de l'estat i evolució de les quasispecies que les mètriques tradicionals de diversitat, les quals simplifiquen estructures de poblacions virals complexes amb un únic valor numèric.

Mitjançant l'aplicació de la categorització QFF a dos pacients amb infecció crònica persistent per VHE que havien fracassat en diversos cicles de tractament amb RBV, hem identificat una discrepància significativa entre l'evolució genotípica i fenotípica quan el virus és sotmès a la pressió de la RBV. L'anàlisi longitudinal va revelar que mentre les poblacions virals es tornaven altament desestructurades a nivell de nucleòtids (genotípic) amb el genotip *master* representat per davall del 10%, es mantenien sorprenentment estables a nivell d'aminoàcids (fenotípic), amb el fenotip *master* representat

entre el 60-80% de la població viral. Aquests resultats suggereixen que el VHE podria aprofitar les mutacions sinònimes per a preservar la funcionalitat proteica i la *fitness* viral, aprofitant l'efecte mutagènic del fàrmac com a catalitzador evolutiu, en compte d'evolucionar envers a una mutagènesi letal o "catàstrofe d'error" caracteritzada per la pèrdua d'informació genètica viral. Aquesta heterogeneïtat genètica augmentada permet al virus adaptar-se més ràpidament a les pressions selectives.

Per complementar els estudis de quasispecies, vàrem aplicar la metodologia de seqüenciació metagenòmica *shotgun* utilitzant mostres seqüencials del pacient del primer estudi, perquè havia mostrat un possible perfil de de resistència a la RBV després d'una segona exposició al fàrmac. Aquest enfocament va permetre identificar diferents substitucions associades a resistència dins del domini aminoacídic de la RdRp, localitzat en la regió genòmica ORF1, incloent-hi D1384N i Y1587F. De manera rellevant, algunes d'aquestes mutacions eren detectables com a variants minoritàries a nivells sub-consens abans de convertir-se en dominants dins de la població o de contribuir al fracàs clínic del tractament. L'estudi també va confirmar l'increment de la diversitat observat en l'anàlisi QFF, així com l'augment de la ràtio de transicions/transversions (Ts/Tv), que actua com una signatura genòmica de l'efecte mutagènic de la RBV.

Des d'un punt de vista clínic, els nostres resultats suggereixen que la monoteràpia amb RBV, especialment en pacients amb càrregues virals elevades de VHE, podria afavorir la selecció de mutacions que incrementen la *fitness* viral. Per tant, les teràpies mutagèniques haurien de combinar-se idealment amb altres inhibidors antivirals capaços de reduir la càrrega viral i per tant el núvol de mutants, minimitzant la probabilitat per a la generació i selecció de resistències. A més, per a previndre les recaigudes virals, el tractament amb RBV s'hauria de mantindre durant diversos mesos fins que l'ARN del VHE siga indetectable en sèrum fent servir l'assaig més sensible disponible, si això fora possible.

Finalment, aquest treball de tesi reforça la importància clínica de la caracterització profunda de variants mitjançant seqüenciació de nova generació (NGS) com a eina predictiva precoç. El monitoratge de variants minoritàries a nivell sub-consens permet anticipar el fracàs terapèutic amb antelació, facilitant la interrupció primerenca de tractaments ineficaços i la implementació d'estratègies de medicina de precisió. La integració del mètode QFF amb l'anàlisi del genoma complet proporciona un marc sòlid per a comprendre la persistència viral i millorar els resultats terapèutics en la infecció crònica per VHE.

SUMMARY

La hepatitis E es una enfermedad inflamatoria del hígado causada por el virus de la hepatitis E (VHE), una de las principales causas de hepatitis vírica aguda en el mundo, con una estimación de 20 millones de infecciones anuales. Aunque la mayoría de las infecciones son autolimitadas, la enfermedad puede evolucionar hacia un estado crónico en pacientes inmunodeprimidos, como los receptores de trasplante de órganos sólidos, y puede conducir a una fibrosis hepática progresiva y a cirrosis. Las guías clínicas actuales recomiendan un tratamiento de tres meses con ribavirina (RBV) en monoterapia. Sin embargo, el fracaso terapéutico primario y la recaída viral continúan siendo problemas importantes, especialmente debido a la falta de alternativas antivirales aprobadas.

En esta tesis se investiga el VHE como un virus estructurado en cuasiespecies, es decir, una población dinámica de variantes estrechamente relacionadas pero genéticamente diferentes que existen como una “nube de mutantes”, principalmente debido a la ausencia de mecanismos de lectura y corrección de errores en la ARN polimerasa dependiente de ARN del virus (RdRp). Para comprender y visualizar mejor estas dinámicas, hemos establecido un nuevo enfoque analítico denominado *Quasispecies Fitness Fractions* (QFF), que estratifica la población viral en cuatro categorías según la frecuencia de los haplotipos: *master*, emergentes, minoritarios y muy minoritarios. Este método proporciona una visualización más completa del estado y la evolución de las cuasiespecies que las métricas tradicionales de diversidad, las cuales simplifican estructuras complejas de poblaciones virales en un único valor numérico.

Mediante la aplicación de la categorización QFF a dos pacientes con infección crónica persistente por VHE que habían fracasado en varios ciclos de tratamiento con RBV, hemos identificado una discrepancia significativa entre la evolución genotípica y fenotípica cuando el virus es sometido a la presión de la RBV. El análisis longitudinal reveló que, mientras las poblaciones virales se volvían altamente desestructuradas a nivel de

nucleótidos (genotípico), con el genotipo master representado por debajo del 10 %, se mantenían sorprendentemente estables a nivel de aminoácidos (fenotípico), con el fenotipo master representando entre el 60 y el 80 % de la población viral. Estos resultados sugieren que el VHE podría aprovechar las mutaciones sinónimas para preservar la funcionalidad proteica y el *fitness* viral, utilizando el efecto mutagénico del fármaco como catalizador evolutivo, en lugar de evolucionar hacia una mutagénesis letal o "catástrofe de error" caracterizada por la pérdida de información genética viral. Esta heterogeneidad genética aumentada permite al virus adaptarse más rápidamente a las presiones selectivas.

Para complementar los estudios de cuasiespecies, aplicamos la metodología de secuenciación metagenómica *shotgun* utilizando muestras secuenciales del paciente del primer estudio, ya que había mostrado un posible perfil de desarrollo de resistencia a la RBV tras una segunda exposición al fármaco. Este enfoque permitió identificar diferentes sustituciones asociadas a resistencia dentro del dominio aminoacídico de la RdRp, localizado en la región genómica ORF1, incluyendo D1384N y Y1587F. De forma relevante, algunas de estas mutaciones eran detectables como variantes minoritarias a niveles sub-consenso antes de convertirse en dominantes dentro de la población o de contribuir al fracaso clínico del tratamiento. El estudio también confirmó el incremento de la diversidad observado en el análisis QFF, así como el aumento de la ratio de transiciones/transversiones (Ts/Tv), que actúa como una firma genómica del efecto mutagénico de la RBV.

Desde un punto de vista clínico, nuestros resultados sugieren que la monoterapia con RBV, especialmente en pacientes con cargas virales elevadas de VHE, podría favorecer la selección de mutaciones que incrementan el *fitness* viral. Por tanto, las terapias mutagénicas deberían combinarse idealmente con otros inhibidores antivirales capaces de reducir la carga viral y, por tanto, la nube de mutantes, minimizando la probabilidad de generación y selección de resistencias. Además, para prevenir las

recaídas virales, el tratamiento con RBV debería mantenerse durante varios meses hasta que el ARN del VHE sea indetectable en suero utilizando el ensayo más sensible disponible, si ello fuera posible.

Finalmente, este trabajo de tesis refuerza la importancia clínica de la caracterización profunda de variantes mediante secuenciación de nueva generación (NGS) como herramienta predictiva precoz. El seguimiento de variantes minoritarias a nivel sub-consenso permite anticipar el fracaso terapéutico con antelación, facilitando la interrupción temprana de tratamientos ineficaces y la implementación de estrategias de medicina de precisión. La integración del método QFF con el análisis del genoma completo proporciona un marco sólido para comprender la persistencia viral y mejorar los resultados terapéuticos en la infección crónica por VHE.

1. INTRODUCTION

1.1 Hepatitis E Virus

1.1.1 HEV Discovery

The Golden era of viral hepatitis research during the 1960s and 1970s enabled the discovery and characterisation of both Hepatitis B Virus (HBV) and Hepatitis A Virus (HAV), ultimately leading to the development of their specific serologic assays[1–4]. However, several burden of acute were occurring in several parts of the world, particularly in India. In 1978, Dr. Mohammad S. Khuroo investigated 269 cases of acute hepatitis in the Baramulla district of the Kashmir Valley (India)[5]. He observed that the outbreak was confined to villages supplied by water from the Ningli Nallah river, strongly suggesting a waterborne aetiology. In parallel, Khorshed M. Pavri retrospectively analysed stored plasma samples from two major epidemics of severe jaundice: one in Delhi in 1955 (29000 cases) and another in Ahmedabad in 1975 (2600) cases[6]. Together, these studies demonstrated that the causative agent tested negative for both A and B antigens and was transmitted enterically, leading to its designation as enterically transmitted non-A, non-B hepatitis (ET-NANBH)[7].

A decisive step in characterising ET-NANBH occurred when Dr. Mikhail S. Balayan, a Russian virologist serving for the Soviet military in Afghanistan, experimentally self-inoculated by ingesting a stool extract coming from an outbreak between the soldiers. He returned to Moscow and, after developing acute hepatitis, electron microscopy of his samples revealed abundant small, non-enveloped, 27–30 nm icosahedral viral particles, providing the first direct visualization of the agent responsible for ET-NANBH in 1983[8]. Genomic characterisation came eight years later, when Dr. Tam from the Genelabs group cloned and sequenced *Burma* strain[9], establishing Hepatitis E Virus (HEV) within the human hepatitis viruses.

1.1.2 HEV epidemiology and genotypes

Hepatitis E is the inflammatory liver disease caused by the infection with HEV. It is one of the leading causes of acute human viral hepatitis worldwide, with a global distribution that reflects complex interactions between viral genotypes, environmental conditions, and socioeconomic factors[10]. Although it has been associated with large waterborne outbreaks previously addressed in this section, HEV is now understood to cause burden across both developing and industrialized regions, driven by diverse transmission pathways including zoonotic spread and foodborne exposure. The World Health Organisation has estimated that it affects 20 million people annually, contributing to considerable morbidity and mortality, causing the death of 44000 patients[11]. Despite its global impact, the real epidemiologic impact of HEV remains underestimated due to limited surveillance specially in low- and middle-income countries[12–14].

The epidemiology of Hepatitis E is shaped by the genetic diversity of the virus, comprising eight described genotypes (HEV1 to HEV8), from which 1 to 4 are most clearly associated with human disease[15,16]. HEV1 and HEV2 are strictly human pathogens that predominate in low- and developing countries, where they are responsible of outbreaks driven by inadequate sanitation, contaminated drinking water and high population density[17]. These genotypes, and particularly HEV1, affect younger adults and carry a high risk of acute liver failure (ALF) in pregnant women, contributing to HEV-associated mortality in endemic regions. On the other hand, HEV3 and HEV4 are considered zoonotic, primarily circulating in pigs, wild boar and deer, causing sporadic infections in high-income countries. In this case, the transmission is mainly due to the consumption of undercooked pork or direct contact with infected animals [18–21], but also after transfusion of blood derivatives[22–25].

The differential distribution of these genotypes is showed in **Figure 1**. Finally, HEV-5 and HEV-6 have only been isolated from wild boar in Japan[26] , HEV-

7 in dromedary camels[27] with one report of human infection by camel meat and milk consumption[28], while HEV-8 was proposed for viral variants isolated from Bactrian camels in China[29].

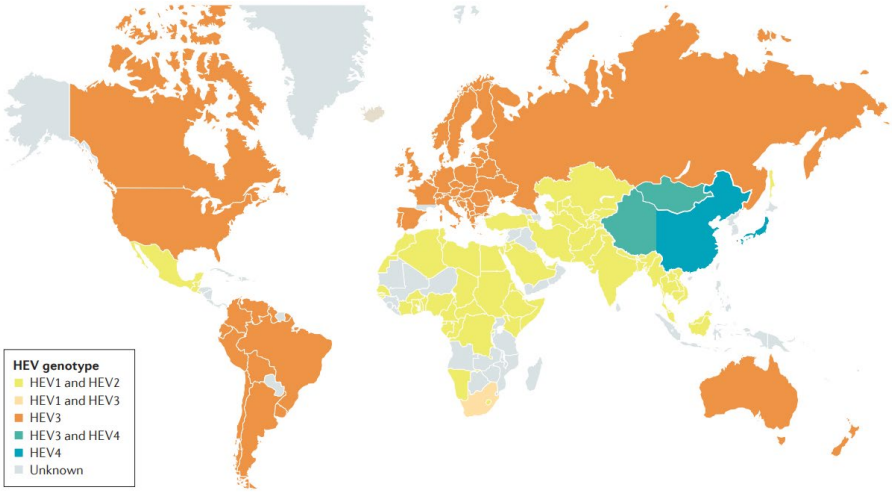


Figure 1. Global HEV distribution of genotypes infecting humans HEV1-4. Differences in genotype distribution are marked by different colours[17].

1.1.3 Molecular Biology of HEV

1.1.3.1 HEV Taxonomy

Hepatitis E virus was the first virus belonging to the *Hepeviridae* family (**Figure 2**) to be described. It includes two subfamilies, being *Parahepevirinae* only detected in fish. On the other hand, *Orthohepevirinae* comprises four genus and nine species, affecting mammals and birds[30]. Commonly, the species related to the different genus are referred to their host, namely *Avihepevirus* as avian HEV[31] while *Rocahepevirus* is usually called rat HEV[32]. The eight HEV1 to HEV8 are comprised inside the recently named species *Paslahepevirus balayani*.

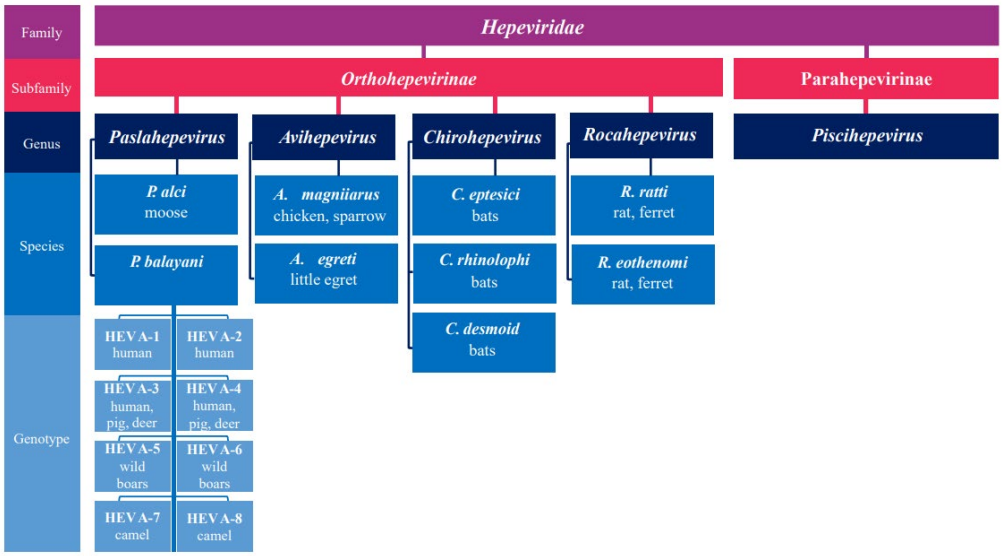


Figure 2. Taxonomic classification of the Hepeviridae family, including Paslahepevirus balayani (HEV)[33].

1.1.3.2 HEV viral particle

HEV produces two distinct types of viral particles during infection: non-enveloped (neHEV) and quasi-enveloped (eHEV) virions[34]. Non-enveloped particles are released in faeces and are icosahedral with a diameter of 27-34 nm and a density approximately 1.25 x cm^{-3}. Virions found mainly in the bloodstream and in the supernatant of HEV-infected cell cultures are quasi-enveloped, surrounded by host-derived lipid membrane acquired by hijacking the cellular exosomal pathway during their release[35]. The eHEV particles, also termed quasi-enveloped HEV particles[34] are more infectious than neHEV particles[36,37], and they are protected from neutralizing antibodies against the viral capsid, performing an important role in immune system evasion[36]. However, neHEV particles have been detected in blood of infected donors, which pose an additional risk to blood transfusion safety, since these particles show different behaviour to pathogen inactivation treatments[38].

1.1.3.3 HEV genome

HEV genome consists of a single-stranded positive-sense RNA molecule of ~7.2 kb in length[9]. It is capped at the 5' end and polyadenylated at the 3' end, allowing the viral RNA to function directly as messenger RNA upon entry into the host cell. The genome contains three major open reading frames (ORFs), although a fourth ORF has been recently described in HEV1[39] (**Figure 3**).

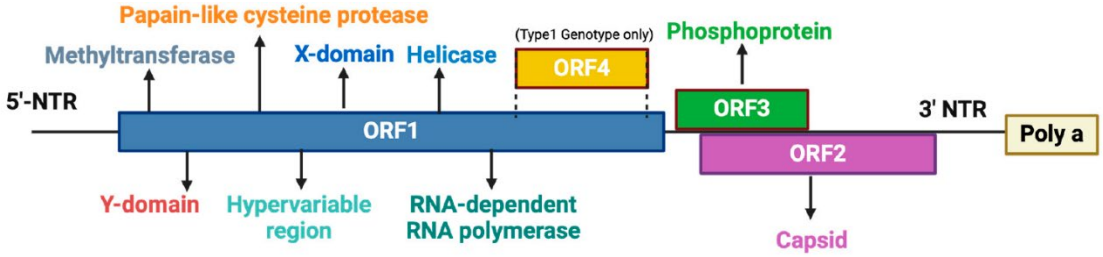


Figure 3. Genomic structure of the HEV. Adapted from Shahini et al.[40] NTR accounts for Non-translated Region.

ORF1 occupies roughly two-thirds of the genome (~5 kb) and encodes a large non-structural polyprotein (pORF1) involved in viral replication[40]. The N-terminal region of pORF1 contains a methyltransferase domain (Met) responsible for the capping of the 5' end of the viral RNA, allowing for RNA stability and efficient translation[41]. Downstream there is the Y domain, highly conserved but poorly characterized. It is thought to contribute to replication complex formation and may play a role in membrane association or protein-protein interactions[42]. Continuously, the Papain-like cysteine protease (PCP) domain exhibits a controversial function[43], since it has been proposed to act as a real proteolytic enzyme for pORF1 and pORF2[44] but also performs multi-vesicular body attachment functions[45]. Hypervariable region (HVR) shows an elevated sequence variability among HEV genotypes, tolerating insertions and deletions. It is associated with host adaptation, cell tropism and enhanced replication efficiency[46]. Following HVR, X domain may have an ADP-ribose binding function[47] involved in counteracting host antiviral responses through ADP-ribosylation, contributing to immune evasion[48]. In addition, it also

interacts with Met and pORF3[49]. The last two domains of pORF1 correspond to Helicase (Hel) and RNA-dependent RNA polymerase (RdRp). Hel domain show NTPase and RNA helicase activities essential for unwinding RNA secondary structures during genome replication while facilitating the mobility of the replication complex along the viral RNA template. RdRp, located at the C-terminus, is the responsible for genomic and subgenomic viral RNA synthesis.

ORF2 encodes the viral capsid protein, responsible for virion assembly, genome packaging and host cell attachment. The ORF2 translation comes mainly from a subgenomic RNA of ~2.2 kb that carries both ORF2 and ORF3[50]. The HEV capsid protein plays a central role in viral assembly and host cell attachment. It is composed by three different protein domains: the shell domain (S), which forms the internal scaffold; the middle (M) domain, which lies on the particle surface; and the protruding (P) domain, similar to spike-like structures involved in host cell binding and the key target for neutralizing immune responses[51,52]. Although the HEV capsid protein contains potential N-glycosilation sites, structural studies indicate that glycosylation is not necessary for virion assembly[53,54].

ORF3 partially overlaps with ORF2. It encodes a small multifunctional protein (VP13) involved in viral budding and egress, intracellular trafficking and modulation of host signalling pathways. It interacts with the cell vesicle machinery mechanisms for the release of eHEV[55,56]. Finally, ORF4 encodes for a small protein that enhances viral replication in HEV1 by interacting with Hel, RdRp and Hel[39].

The genomic organization of HEV is highly conserved across genotypes. However, there is intense sequence variation that contributes to differences in viral host range, transmission routes and disease routes, among others[57].

1.1.3.4 HEV viral cycle

1.1.3.4.1 Viral cell entry and uncoating

Hepatitis E virus (HEV) is primarily transmitted via the faecal–oral route. However, the mechanisms by which the virus crosses the intestinal barrier and enters to the portal circulation remain incompletely understood. Experimental evidence using animal models and human-derived intestinal cell culturing has demonstrated that HEV can infect and replicate in intestinal epithelial cells, supporting the intestine as an initial site of infection[58–60]. Nevertheless, intestinal replication role in HEV spread across the whole body has not been definitively established. Current data suggest that the first productive infection within the host is mediated predominantly by quasi-enveloped HEV (eHEV), while non-enveloped HEV (neHEV) may contribute under specific conditions.

Following intestinal replication, HEV is thought to cross the intestinal vascular endothelium and reach the liver via the portal venous system, where hepatocytes are infected from the basolateral surface[58]. Infectious virions are released primarily into the bloodstream as quasi-enveloped particles, resulting in viremia driven by eHEV, mainly egressing apically, whereas a minor fraction is released basolaterally[61]. A smaller proportion of non-enveloped particles has also been detected in circulation[38] particularly during later stages of infection where there is liver damage, although their precise contribution to intra-host spread remains unclear[59].

The existence of both non-enveloped and quasi-enveloped HEV particles has important implications for viral entry. neHEV enters cells through clathrin-mediated endocytosis and interacts with heparan sulfate proteoglycans and additional host factors[62–64]. In contrast, eHEV is internalized via endocytosis and requires trafficking through the endolysosomal system to remove its host-derived membrane prior to genome release (**Figure 4**). Several cellular

factors, including TIM-1, GRP78, ATP5B, ASGPR, integrin α3, and EGFR, have been implicated in HEV attachment or entry[65–69]. Nevertheless, none has been conclusively identified as a primary receptor, and their relative importance may vary depending on cell type and viral particle form.

Endosomal transport plays a particularly important role in eHEV infection. Rab5- and Rab7-dependent trafficking, as well as lysosomal lipid degradation and cholesterol processing, are required for productive eHEV entry, suggesting that removal of the quasi-envelope occurs within late endosomal or lysosomal compartments[59,64]. Inhibition of endosomal acidification or lipid degradation[64] markedly reduces eHEV infectivity, whereas neHEV entry is largely unaffected by these interventions[59]. Despite these advances, the precise molecular triggers that lead to capsid uncoating and genome release into the cytoplasm remain poorly defined.

1.1.3.4.2 Viral assembly and release

Following entry and uncoating, the HEV genomic RNA serves as a template for translation of the ORF1 polyprotein, which drives viral RNA replication through the synthesis of negative-strand intermediates, full-length genomic RNA and subgenomic RNA. ORF1 encodes for the replication-associated function' polyprotein, but its proteolytic processing during natural infection remains controversial[43,70,71]. Accumulating evidence suggests that ORF1 lacks intrinsic protease activity and that any processing observed is likely mediated by host proteases, while others assign the proteolytic function of the PCP Domain[72].

Next, structural protein synthesis occurs via translation of the subgenomic RNA, producing ORF2 and ORF3[50]. Virion assembly involves encapsidation of genomic RNA by the non-glycosylated form of the ORF2 capsid protein, which may occur in specialized replication-associated membranes, although

the precise intracellular site of assembly has not been conclusively identified. In contrast, glycosylated and secreted forms of ORF2 do not contribute to virion formation and are thought to act as immune decoys[54,73,74]. ORF2 also displays complex intracellular trafficking, including transient nuclear localization early during infection.

ORF3 is not required for capsid assembly but is essential for viral egress and the formation of quasi-enveloped particles. HEV employs the host endosomal sorting complex required for transport (ESCRT) machinery to bud into multivesicular bodies, generating exosome-like vesicles that contain the viral capsid and ORF3 protein[75]. A conserved PSAP motif within ORF3 mediates interaction with ESCRT components such as Tsg101, enabling virion release[76,77]. Palmitoylation of ORF3 further promotes membrane association and efficient secretion[78]. Although this ESCRT-dependent pathway is well supported experimentally, the full range of host factors involved in eHEV biogenesis in vivo remains to be defined.

Following release, eHEV particles circulate in the bloodstream, where their membrane confers resistance to neutralizing antibodies. Upon entering through the biliary tract, bile acids disrupt the quasi-envelope, resulting in the release of stable, non-enveloped virions into the faeces. This dual lifestyle allows HEV to evade host immunity during intra-host spread while keeping environmental stability for faecal–oral transmission. Beyond hepatic infection, new studies evidence suggests that quasi-enveloped HEV may contribute to extrahepatic dissemination, including potential neuroinvasion[79]. **Figure 4** resumes the viral HEV cycle.

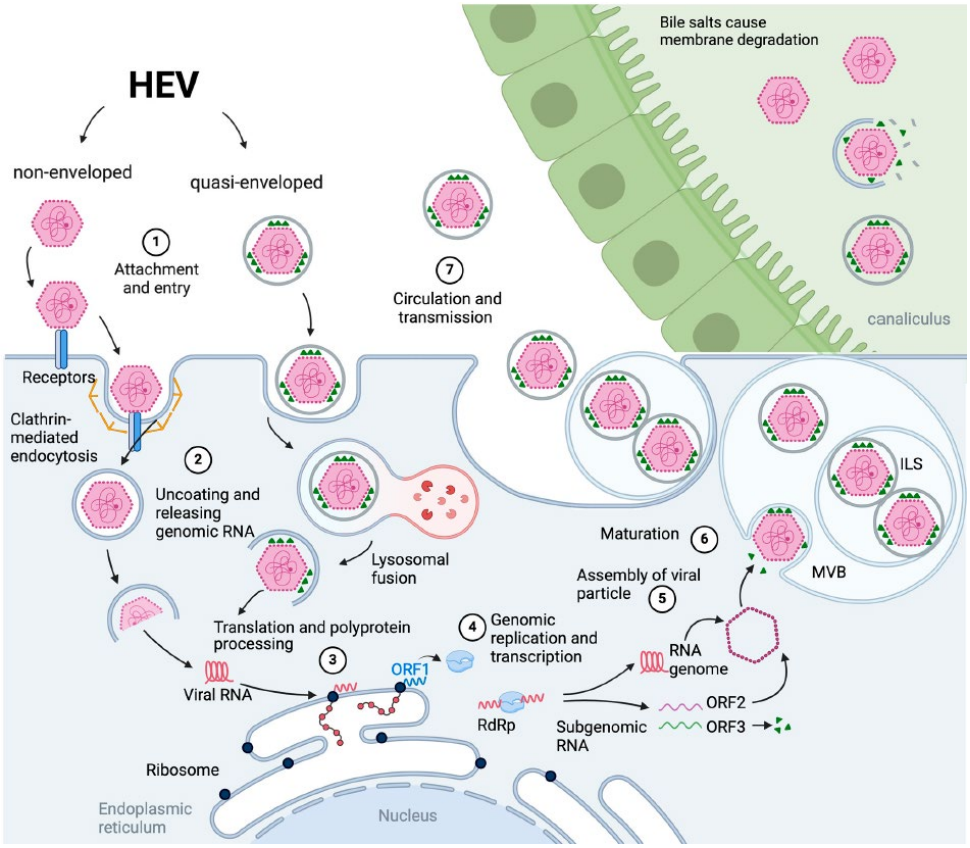


Figure 4. HEV viral cycle for both eHEV and neHEV[40].

1.2. Clinical Features of HEV Infection

1.2.1 Acute infection

HEV infection begins with an incubation period ranging from 3 to 8 weeks, with an average duration of approximately 5-6 weeks, after which acute disease may develop[80]. Viral replication becomes detectable in the liver seven days after the initial infection[81], leading to the release of newly formed virions excreted predominantly via the biliary tract into the faeces, although it has also been detected for urine excretion. The acute phase of the infection typically lasts between 1 to 6 weeks and is followed by a convalescent phase characterized by gradual clinical and biochemical recovery. HEV infection is typically self-limited, being resolved through host immune responses in immunocompetent patients.

In most cases, HEV infection is asymptomatic or results in mild, non-specific systemic illness without clear evidence of hepatic injury. However, approximately 5-30% of infected individuals develop acute icteric hepatitis. Symptomatic disease generally begins with a short initial phase marked by malaise, low-grade fever, nausea and vomiting, followed by an icteric phase characterized by jaundice and dark-coloured urine[82]. Biochemically, alanine aminotransferase (ALT) levels in plasma rise during these two phases, while bilirubin concentration peak during the icteric phase[83]. Clinical symptoms and biochemical markers usually resolve within the convalescent phase, with no need of antiviral treatment.

Although most acute HEV infections follow a benign course, a small proportion of infected individuals (0.5-4%) progress to acute liver failure (ALF), which is associated with high mortality[84]. The risk of ALF is substantially increased in patients with pre-existing chronic liver disease, with reported mortality rates reaching up to 67%[85].

The clinical spectrum of acute hepatitis E is strongly influenced by viral genotype. Infections caused by HEV1 and HEV2 range from asymptomatic or mild disease to severe acute hepatitis and fulminant liver failure. HEV1 and HEV2 could be also of important severity during pregnancy, particularly in the second and third trimesters. In this population, the risk of progression to ALF is highly increased, with mortality rates from 15 to 25%[86]. Furthermore, they are also associated with another obstetric outcomes, including miscarriage, preterm delivery, stillbirth and perinatal mortality. Vertical transmission from the mother to the foetus may occur, placing neonates at risk of hepatitis and severe neonatal complications[87].

In contrast, infections with HEV3 and HEV4 are predominantly asymptomatic and are frequently identified through incidental detection of elevated liver-related enzymes or seroconversion in population-based studies. Large-scale epidemiological investigations during the Chinese HEV vaccine indicate that

fewer than 2% of HEV3-HEV4 infected patients clinically develop hepatitis[88], although symptomatic disease including jaundice is reported in 5-33% of cases in European cohorts[89,90]. ALF is rare in HEV3 and HEV4 infections, regardless of underlying liver disease, although sporadic cases have been described[91,92]. Importantly, the increased disease severity observed in HEV1 and HEV2 infections is not seen with HEV3 or HEV4.

During the course of the infection, the host starts a humoral immune response characterized by the initial production of IgM, followed by isotype switching to IgG, which establishes long term adaptative immunity (**Figure 5**). Accordingly, the detection of anti-HEV IgM and IgG antibodies constitutes the primary serological markers for infection[93]. Nevertheless, reinfection has been reported in individuals with low anti-HEV IgG titres (<7 IU/mL)[94], although these secondary infections are generally associated with reduced disease severity[95].

Despite the widespread use of serological assays, the detection of HEV RNA by reverse transcription polymerase chain reaction (RT-PCR) remains the gold standard for the diagnosis of active HEV infection[96]. RT-PCR allows direct identification of viral RNA in blood and stool, being particularly valuable during the early phase of infection, before the development of a detectable antibody response, as well as in cases with atypical or absent serological markers. RT-PCR in combination with a established WHO international reference panel permits real RNA quatification[97]. Moreover, molecular detection is essential for confirming ongoing viral replication, distinguishing acute infection from past exposure, and enabling viral genotyping, which has important epidemiological and clinical implications.

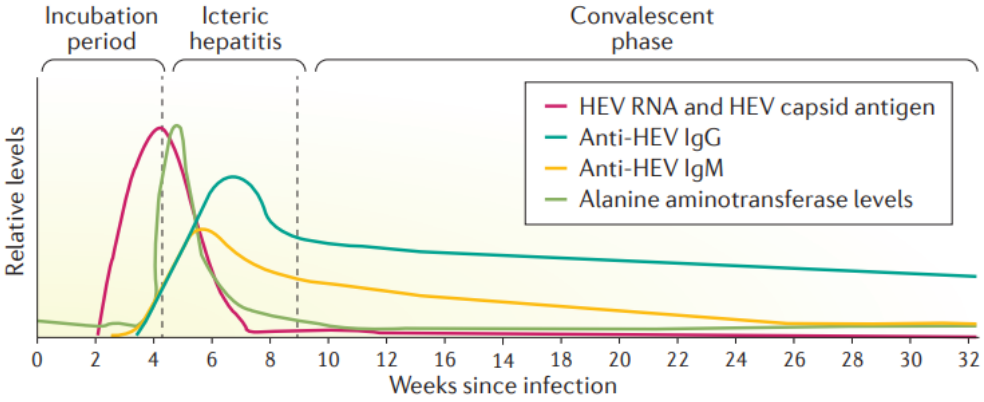


Figure 5. HEV infection course description in weeks and its associated biomarkers[17].

1.2.2 Chronic HEV

HEV can cause chronic infection in immunocompromised patients, with a persistent replication beyond 3-6 months and lead to chronic hepatitis with progressive liver fibrosis and the later cirrhosis[98]. Chronic HEV has been described in patients with impaired immune system including solid organ[99] and hematopoietic stem cell transplant recipients[100], patients undergoing chemotherapy[101] or immunotherapy and patients with HIV infection[102]. This chronic infection clinical course has been reported predominantly for HEV3, but HEV4 has been also associated to chronic Hepatitis E infection[103].

Among immunocompromised patients, solid organ transplant recipients have been the most extensive studied group. Approximately one third of infected transplant recipients spontaneously clear the virus, while the remaining develop chronicity that leads to later cirrhosis in approximately 10% of the cases[104–106]. The probability to develop chronic hepatitis E depends on the intensity of immunosuppression, with more frequent viral persistence in patients treated with tacrolimus compared with cyclosporine A[106], as well as in those with reduced lymphocyte counts and attenuated HEV-specific T cell responses[107].

Clinically, chronic HEV infection diagnosis is further complicated by the limited sensitivity of serological assays, as anti-HEV IgM and IgG may be absent during both acute and chronic infection[106], reinforcing the role of RNA detection for RNA accurate diagnosis.

1.2.3 Extrahepatic manifestations

Although hepatitis E virus primarily infects the liver, HEV infection can be associated with a range of extrahepatic manifestations[108,109]. These manifestations are reported most frequently in HEV genotype 3 infections and highlight the capacity of the virus to affect multiple organ systems.

Neurological complications are the most commonly described extrahepatic manifestations and include Guillain–Barré syndrome[110,111], neuralgic amyotrophy[112,113], meningoencephalitis[114], and peripheral neuropathies. HEV infection has also been associated with renal and haematological disorders, such as membranous and cryoglobulinemic glomerulonephritis[115] or thrombocytopenia[116], haemolytic and aplastic anemia[117,118].

1.2.4 HEV treatment

1.2.4.1 Ribavirin

Ribavirin (RBV, 1-β-D-ribofuranosyl-1,2,4-triazole-3-carboxamide) is a synthetic guanosine nucleoside analogue that displays broad antiviral activity against RNA and DNA viruses in vitro, and has been extensively used as a broad-spectrum antiviral agent for the treatment of hepatitis C (HCV)[119], respiratory syncytial virus (RSV)[120], Lassa fever virus[121], and Crimean-Congo hemorrhagic fever[122].

Despite its well-established antiviral efficacy, ribavirin operates through multiple overlapping mechanisms rather than a single dominant pathway, making its complete mechanism of action a subject of ongoing scientific investigation. The primary mechanism by which ribavirin exerts its antiviral activity involves competitive inhibition of inosine monophosphate dehydrogenase (IMPDH), the rate-limiting enzyme in the de novo synthesis pathway of guanosine nucleotides. Following cellular uptake via nucleoside transporters and phosphorylation to its active triphosphate form (RTP) by adenosine kinase, ribavirin monophosphate mimics inosine-5-monophosphate and competitively binds to the substrate-binding site of IMPDH, thereby depleting intracellular pools of guanosine triphosphate (GTP). Since RNA viruses depend on abundant intracellular GTP for rapid genomic replication, this GTP depletion fundamentally impairs viral protein synthesis and genome replication. A second critical mechanism involves lethal mutagenesis through incorporation of RTP into the nascent viral RNA by the viral RNA-dependent RNA polymerase (NS5B in hepatitis C virus). Ribavirin exhibits ambiguous base-pairing capability, binding both cytosine and uracil, which results in the introduction of random mutations throughout the viral genome at rates exceeding the error threshold tolerable for viral fitness, a phenomenon termed "error catastrophe[123,124]".

This mutagenic mechanism drives the production of non-viable, low-fitness viral quasispecies, effectively eliminating the viral population through accumulated genetic damage rather than direct polymerase inhibition, characterized by G-to-A and C-to-U nucleotide transitions[119].

Additionally, RBV functions as a weak inhibitor of viral polymerase through direct competitive inhibition of nucleotide binding, though this mechanism likely contributes more substantially in contexts of low viral load where interferon-mediated viral suppression has created favourable conditions for RBV's activity. Beyond its direct antiviral mechanisms, RBV exerts significant immunomodulatory effects by orchestrating a shift in the T helper cell balance

toward Th1-dominated immunity, and more effective antiviral effect. RBV has demonstrated the capacity to directly induce interferon-stimulated genes at modest levels and may potentiate the interferon response through recognition by toll-like receptors or inhibition of transcriptional repressors, thereby amplifying the innate antiviral state within infected tissues[124].

This multifaceted mechanism of action has proven resilient against the development of viral resistance, particularly when RBV is employed in combination with interferon or direct-acting antiviral agents, establishing it as a cornerstone therapeutic agent in antiviral medicine despite the ongoing scientific quest to definitively elucidate which mechanism provides the dominant contribution to its broad-spectrum antiviral activity.

However, in some viral infections including HEV chronic infection, combination therapy is not a possibility since direct-acting antiviral are, at present, not available.

Current clinical guidelines including the European Association for the Study of the Liver (EASL) advocate for an initial reduction of the immunosuppressive therapy as primary intervention for chronic HEV infection, particularly in solid organ transplant recipients (**Figure 6**)[125]. When viral clearance is no achieved, off-labelled RBV monotherapy is administered for three months, using either weight-based dosing or adjustment according to renal function. Sustained virologic response (SVR) rates up to 78% have been reported[126,127]. In cases of viral relapse or primary treatment failure, an extended second course of RBV, typically lasting six months, is recommended. Despite this, for patients with no achieving SVR with the second RBV treatment, there is no other option rather than pegylated interferon-α for liver transplant recipients.

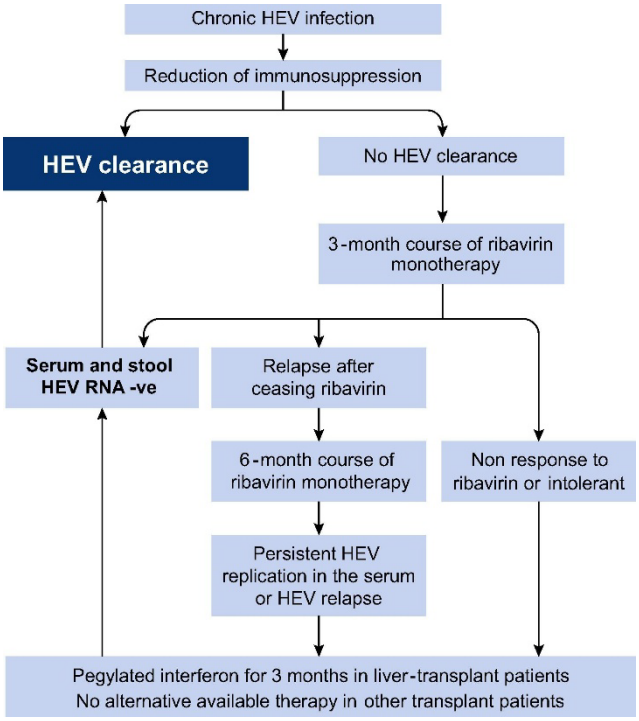


Figure 6. EASL clinical guidelines for chronic HEV infection management[125].

A major limitation of RBV monotherapy in HEV is dose-limiting hemolytic anemia, which frequently necessitates dose reduction and compromises antiviral efficacy[127]. The absence of approved alternative antivirals and the contraindication of RBV in pregnant women due to its teratogenicity further underscore the unmet clinical need for safer and more effective therapeutic strategies against chronic HEV infection[83].

Despite its clinical efficacy, the precise mechanism by which RBV mediates HEV clearance remains incompletely understood. Proposed mechanisms include depletion of intracellular guanosine triphosphate (GTP) pools[128] and the induction of error-prone viral replication leading to lethal mutagenesis[129,130]. Work by Marion et al. revealed that RBV has differential effect on HEV release from intestinal epithelial cells with a pronounced inhibition of viral liberation at the basolateral membrane in vitro[60]. Conversely, apical surface presented more limited effect in HEV viral egress. This highlights the importance of the study not only liver but gut HEV replication

when evaluating antiviral efficacy. In addition, evidence suggests that the intestinal compartment may serve as reservoir for HEV populations with reduced sensitivity to RBV, potentially contributing to relapse after treatment cease[60]. Recent data show that HEV can entry, replicate and propagate in human kidney cell lines and that viruses derived from urine were more distinct from those derived from plasma and stool, and showed less RBV sensitivity in inhibiting HEV replication in some kidney cells, suggesting that kidney cells may represent a new reservoir for patient relapse[131].

Finally, several RBV resistance associated mutations for HEV have been described in previous studies using monotherapeutic regimens with RBV. Interestingly, most of them are related to RdRp in pORF1, including HVR insertions, Y1320H, K1383N, K1398R, V1479I, Y1587F and G1634R/K[132–134].

1.2.4.2 Alternatives to ribavirin

Pegylated interferon-α has showed antiviral activity against HEV in liver transplant recipients, however, its use is contraindicated in other transplants due to its associated risk of graft rejection[125].

As it happened with RBV, with an off-labelled use in HEV but widely applied for Hepatitis C virus (HCV), Respiratory Syncytial Virus (RSV) or viral haemorrhagic fever treatments, other previously approved antivirals have been tested to deal with HEV infection. Alternative therapeutic strategies, including combination therapy of RBV with the polymerase inhibitor sofosbuvir have been intensely explored yielding positive but limited outcomes[135–137]. What is more, HEV resistance-associated variants during sofosbuvir treatment have been also described[138], specifically amino acid substitution A1343V in the RdRp. In addition, combination of RBV and sofosbuvir resistance amino acid substitutions have been detected in patients

with an initial failed RBV treatment followed by combination with sofosbuvir [139], indicating the need to design for novel next-generation antivirals[140].

Netzler et al. performed a screening study of anti-HEV compounds, testing 16 nucleoside and non-nucleoside antivirals including from developmental compounds to FDA-approved drugs. Combination of GPC-N114 and NITD008 were identified as the most potent HEV replication inhibitors in cell based assays[141]. Furthermore, NITD008 in combination with sofosbuvir has also showed inhibition of HEV replication in hepatic cells but showed reduced efficacy in HEV infected neuronal cells[142]. Other nucleoside analogue, favipiravir, has showed anti-HEV effect by inhibiting RdRp activity, with additive antiviral effects when used in combination with sofosbuvir[143].

Alternatively, zinc salts have showed inhibition of the RdRp with decrease in viral replication *in vitro*[144]. However, no reduction in viral titters have been observed in chronic patients which underwent transplant although presenting higher intra-erythrocyte zinc levels[145]. Several plant-based molecules including silvestrol[146], rocaglamides[147] and isocotoin[148] have also shown efficacy in HEV replication reduction *in vitro*, but there is limited *in vivo* evidence.

1.3. Viral Quasispecies and dynamics

1.3.1 Error-prone replication and quasispecies dynamics

RNA virus genetics is intrinsically related to the error rate occurring during viral template replication. The median mutation rates for RNA viruses range from 10^{-3} to 10^{-6} mutations per nucleotides incorporated and per cycle of replication[149]. This is the first step in the diversification of viruses in nature. In some viruses, such as coronaviruses, the polymerase incorporates a 3′→5′ exonuclease proofreading domain that partially compensates the higher error rate. However, most RNA viruses lack the proof-reading activity, for example in Hepatitis C virus, HIV, Influenza, Poliovirus, Dengue or Zika virus[150].

The errors introduced increase the variability of the population in each replication cycle. Hence, the viral populations exist not as genetically uniform entities but as dynamic and diverse ensembles of closely related variants, called as quasispecies[123,151]. This genetic heterogeneity allows rapid adaptation to selection pressures including immune responses and antiviral therapies, that could lead to the emergence of drug-resistant variants through fitness selection. The mutant cloud is then always evolving, and its size depends on the viral load and replication capacity[150]. Each one of the unique genomes is called an haplotype. Environmental changes drive the success or elimination of each haplotype within the quasispecies, a process governed by the differential fitness of viral variants under selective pressure. Higher fitness implies that the haplotype frequency will increase during the evolution of the quasispecies, while lower fitness may imply a decrease in the haplotype frequency throughout the viral infection.

As an example, **Figure 7** illustrates the view of a quasispecies at three consecutive different moments of viral population evolution, in which the asterisk represents a higher fitness mutation that gets selected until it appears in the consensus sequence in the last point of the diagram.

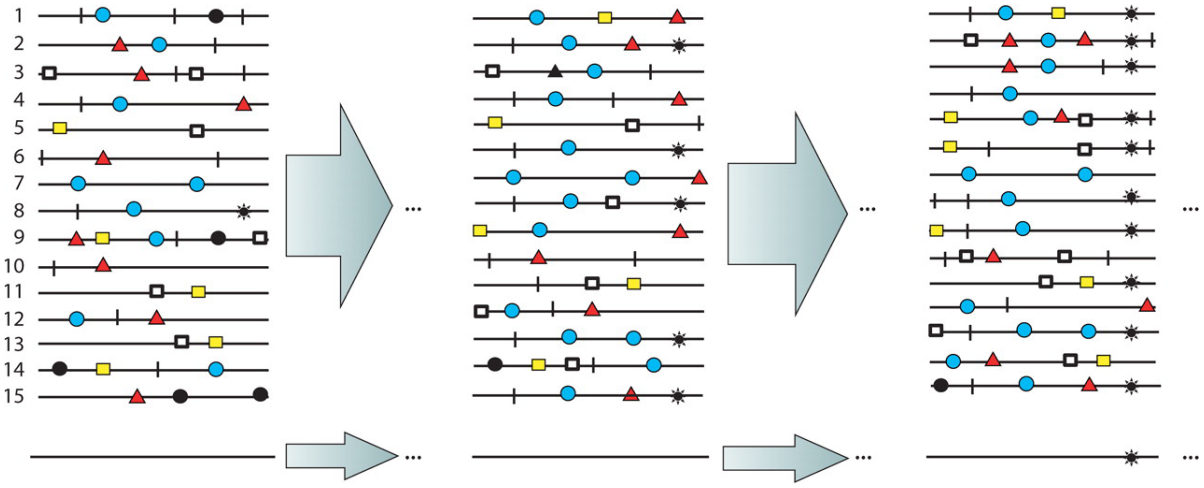


Figure 7. Quasispecies evolution represented at three different points. Each row (1 to 15) represents a different haplotype. The line in the bottom corresponds to the consensus sequence. Adapted from Domingo et al.[152]

However, this accumulation of diversity is not exempt of limitations. The error threshold or lethal mutagenesis theory stated that there is a limit of mutations that a viral population can tolerate[153–155]. This represents the maximum mutation rate at which genetic information can be stably maintained, to avoid the loss of genomic integrity that leads to what is called error catastrophe. Under these conditions, the viral population lose its capacity for sustained viral replication. This concept has important implications for antiviral strategies that aim to drive viruses to the error threshold to achieve viral extinction[156].

1.3.2 Methods to study viral quasispecies

Research in viral quasispecies has transitioned from low-resolution techniques to high-throughput sequencing and advanced computational analyses that enable detailed characterization of intra-host viral populations[157].

Historically, viral quasispecies were examined through molecular cloning of viral sample followed by Sanger sequencing of individual clones[158]. Although this method provided high base accuracy, it was limited in throughput and often failed to capture rare variants that may contribute to key factors such as drug resistance[159]. Nevertheless, these approaches enabled the initial characterization of complex viral populations.

The development of next generation sequencing (NGS) and the application of deep sequencing has allowed to increase the depth at which viral populations have been addressed, mainly through amplicon-based studies[160]. This has enabled the detection of hundreds to thousands of haplotypes and quantify their relative abundances, including low frequency haplotype variants.

Following sequencing, all reads must be aligned using multiple sequence alignment algorithms. Secondly, equal sequences are collapsed into haplotypes, and its frequency calculated. After that, several diversity metrics have been developed to assess viral complexity, including statistical indexes previously used in ecology[152,161].

The most commonly used indices include the minimum and maximum mutation frequency (Mfmin and Mfmax), the normalised Shannon entropy (H_s) and nucleotide diversity (π). Mfmin is defined as the ratio between the number of positions with a detected mutation and the total number of nucleotides sequenced (number of reads x size of the amplicon). Mfmax, by contrast, represents the proportion of nucleotides carrying mutations within a specific sequence's population[161].

H_s is one of the most widely used metrics for quantifying diversity within a population, integrating both haplotype richness (number of distinct haplotypes) and evenness (relative abundances of each haplotype). When normalised, H_s becomes an index varying from 0 to 1[161,162].

π is a population genetics metric that measures the average number of nucleotide differences per site between any two randomly chosen sequences from a population[163]. Unlike H_s, which focuses on the distribution of variant frequencies, π directly reflects the extent of genetic divergence at the sequence level, capturing cumulative mutational burden. What is more, π is more robust than H_s regarding sample size variations.

Finally, Hill numbers (^qD)[164] provide a unified diversity framework that spans across diverse traditional indexes by adjusting the order q (**Equation 1**).

- q=0 Number of total haplotypes
- q=1 Exponential of Shannon Entropy
- q=2 Inverse of the Simpson index
- q=3 Inverse of the predominant haplotype frequency

$$^qD(p) = \left(\sum_{i=1}^{H} p_i^q\right)^{1/(1-q)}$$

Equation 1. Hill number equation for the haplotype frequencies observed in the quasispecies.

More recently, indices such as Rare Haplotype Load (RHL) have been described to study and quantify the proportion of low-frequency variants in a quasispecies[165], being proposed as sensitive marker of mutagenesis or population destabilization. RHL captures the fraction of reads belonging to haplotypes below a frequency threshold (1%), offering a complementary perspective on quasispecies complexity beyond classical diversity indexes.

1.3.3. Quasispecies application to diagnostics

Viral quasispecies studies have had a profound impact on clinical virology by direct changing how chronic viral infections are monitored and treated. Clinically, this framework has helped to explain viral persistence, immune escape and variable disease progression. For viruses such as Hepatitis B

Virus (HBV), Hepatitis C Virus (HCV) and Human Immunodeficiency Virus (HIV), quasispecies analyses have provided tools to interpret treatment failure and anticipate viral evolution under selective pressure[152,166].

In HCV infection, quasispecies studies have been particularly influential because of the viral high variability and its capacity to establish chronic infection[167]. Measuring intra-host diversity helped to correlate quasispecies complexity with treatment response, disease severity and likelihood of viral clearance.

In our laboratory we had previously translated the quasispecies concept to clinical meaningful insights for HCV[168–170]. Our studies demonstrated that minor variants within HCV quasispecies could determine treatment outcome, long before they become dominant[159]. By using deep sequencing, we showed how baseline quasispecies structure, fitness and resistance-associated substitutions (RAS) influence the response to direct-acting antivirals (DAAs)[169,171]. This helped clinicians to conceive that treatment failure is not only due to new generated mutations but by pre-existing variants selected under therapy.

1.3.4. HEV as a quasispecies virus

HEV presents a high fixation mutation rate (evolution rate) of 1.5 base substitutions per site per year, which is similar to the reported for HCV[172]. The high variability and elevated replication kinetics make HEV population structure compatible with quasispecies.

Several studies have shown evidence that HEV populations behave like quasispecies[173–175]. *Grandadam et al*[176] presented the first HEV clone study using prospective plasma samples from Tanefdour in Algeria taken during the epidemics of 1986. They revealed HEV variability and quasispecies-like structure, explaining the adaptability of the virus during the epidemics. Years

later, *Lhomme et al.* related HEV quasispecies diversification with liver fibrosis progression in chronic patients[177]. Finally, HEV quasispecies have been also detected in three HEV patients manifesting neurological sympthoms[178].

1.4 Summary and Aims

The increasing availability of ultra-deep sequencing technologies has markedly improved the research in RNA viral quasispecies. High sequencing depth offers the potential to explore viral population structure with great detail, supporting the detection of minority variants and the assessment of genetic diversity. Nevertheless, the analytical thresholds commonly applied to ensure accuracy may limit the detection of low-frequency variants that might be biologically relevant components of the viral population. This limitation is particularly important in the context of antiviral therapies with mutagenic activity such as RBV. Consequently, there is a growing need for analytical approaches capable of reliably characterizing these low-frequency variants to better capture the full complexity of viral quasispecies.

In patients with chronic HEV infection who do not respond to RBV monotherapy, a detailed investigation of viral quasispecies may provide insight into mechanisms associated with treatment failure and inform therapeutic decision-making. A thorough molecular characterization of HEV quasispecies dynamics during treatment may be an important component in improving the clinical management of chronic HEV infection.

2. HYPOTHESIS

A novel quasispecies analysis approach based on stratifying viral haplotypes by its frequency will provide a more accurate definition of the molecular status of viral quasispecies than current methods, and will enable the identification of evolutionary patterns linked to therapeutic response and chronic infection persistence. Comprehensive analysis, including minor variants, will reveal the timing of resistance-associated mutation emergence and support their use as early predictors of antiviral (ribavirin for HEV) treatment failure.

3. OBJECTIVES

3.1 General objectives

1. To evaluate and compare haplotypes frequency-based analysis for quasispecies studies across different viral evolution scenarios.
2. To apply the new frequency-based quasispecies analysis to further characterize the chronic HEV infection under RBV pressure.
3. To assess the HEV chronic infection RBV resistance development using *shotgun* whole genome sequencing.

3.2 Specific objectives

Study 1

1. To develop a method to characterize the molecular status of viral populations by stratifying quasispecies haplotypes into proportion-based subsets, enabling inference of quasispecies fitness.
2. To apply this method to different sample sets, including both structured and unstructured quasispecies populations, and compare and complement the results with diversity estimates based on Hill numbers (^{q}D).

Study 2

1. To describe in-host quasispecies evolution of HEV in patients with long-term chronic infection.
2. To analyse and compare population dynamics at the genotypic (nucleotide) and phenotypic (amino acid) levels.
3. To assess the impact of RBV treatment on quasispecies structure.

Study 3

1. To apply shotgun metagenomic sequencing to recover the full-length HEV genome from a chronically infected patient with RBV resistance.
2. To characterize genome-wide variation across sequential RBV treatment courses in order to identify, track, and evaluate the emergence and fixation of RBV resistance-associated mutations.
3. To perform frequency-based deep variant profiling to detect RBV resistance mutations at early stages, and to assess their longitudinal dynamics within the viral quasispecies.

4. COMPENDIUM OF ARTICLES

4.1 Study 1

Quasispecies Fitness Partition to Characterize the Molecular Status of a Viral Population. Negative Effect of Early Ribavirin Discontinuation in a Chronically Infected HEV Patient.

4.1.1 Article 1 Summary

4.1.1.1 Introduction

Viral quasispecies are dynamic populations of closely related genomes (haplotypes) generated during replication by error-prone viral polymerases. The relative frequency of each haplotype in a population serve as indirect measure for its fitness, the replication capacity in competitions with other genomes. Traditionally, quasispecies diversity has been summarized using indices like Shannon entropy or nucleotide diversity, but these often reduce complex structures into a single value. There is the need for methodologies that can visualize changes over time, especially during viral chronic infections.

4.1.1.2 Objectives

1. To develop a method to characterize the molecular status of viral populations by stratifying quasispecies haplotypes into proportion-based subsets, enabling inference of quasispecies fitness.
2. To apply this method to different sample sets, including both structured and unstructured quasispecies populations, and compare and complement the results with diversity estimates based on Hill numbers (^{q}D).

4.1.1.3 Study design

- Three different viral datasets were selected for quasispecies study:
    - Technical clone: a synthetic SARS-CoV-2 spike gene used to establish a baseline for a highly structured population with predefined low-level mutations.
    - Controlled *in vitro* experiment: HCV-infected cell cultures treated with mutagenic agents (RBV and Favipiravir) or an RNA-

dependent RNA-polymerase inhibitor (Sofosbuvir) to evaluate characteristic shifts in quasispecies composition.

- Natural infection, longitudinal follow-up: A 27 year old patient, chronically infected with hepatitis E virus at the time of diagnosis, who failed three RBV treatment regimens, providing 13 longitudinal samples for analysis.

- Sequencing was performed using Illumina MiSeq platform to sequence amplicons of 300-500bp. Deep sequencing was performed to ensure 10^5reads/amplicon. The SARS-CoV2 dataset consists of 12 amplicons, the HCV dataset consists of three amplicons, and the HEV dataset consist of one amplicon.
- Data was processed to maintain full-length read integrity and bootstrap was applied to all diversity calculations to correct for variations in the sample size.

4.1.1.4 Results according to objectives

4.1.1.4.1 Study 1. Objective 1: To develop a method to characterize the molecular status of viral populations by stratifying quasispecies haplotypes into proportion-based subsets, enabling inference of quasispecies fitness.

The availability of ultra-deep sequencing datasets enabled a highly robust characterization of RNA viral quasispecies. In previous studies, it was proposed that at sequencing depths of ≥10,000 reads per sample, haplotypes present above 0.5% abundance could be identified without erroneous mutations and with high confidence. This approach also allowed reliable detection of resistant mutants as minor variants at ~1% frequency and provided accurate estimates of quasispecies complexity[168]. However, it was already evident that this conservative threshold inevitably led to the loss of a substantial fraction of real viral variants present below 0.5%, which nonetheless belonged to the true viral population. This limitation became particularly critical when treating patients with RBV or any other drug with mutagenic properties, where a large number of haplotypes are expected to fall below this abundance threshold. To overcome this constraint and enable the analysis of low-frequency but biologically meaningful variants, new analytical strategies were required. This led us to the development and application of a method based on haplotype frequencies, called quasispecies fitness fractions (QFF). The method divides the quasispecies in four fractions:

- Master: the fraction of molecules belonging to the most frequent haplotype; that is, the one present at the highest percentage, representing the sequence with higher fitness of the quasispecies.
- Emerging: the fraction of molecules with a frequency higher than 1% and less than the master percentage belonging to haplotypes that are able to compete with the predominant one, but having the possibly to replace it.

- Low fitness: the fraction of molecules at frequencies from 1% to 0.1%, belonging to haplotypes that have a low probability of progressing to higher frequencies, but not null.
- Very low fitness: the fraction of molecules with frequencies <0.1% belonging to haplotypes with very low fitness including defective genomes. The likely fate of these molecules individually is degradation, but the fraction is continuously fed with new very low fitness genomes produced by replication errors or by host editing activities, or caused by RBV mutagenic effect.

The representation of the four fractions (**Figure 8**) was done using a histogram plot with each of the fractions represented by a different colour: master (red), emerging (green), low fitness (blue) and very low fitness (purple).

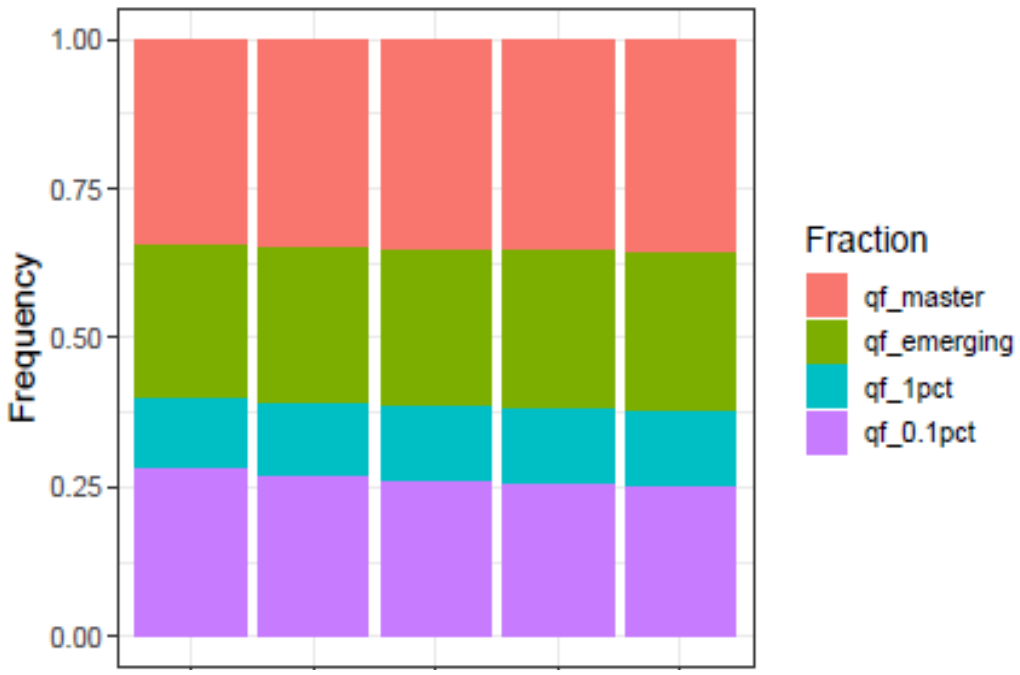


Figure 8. Example figure of the QFF analysis applied to 5 samples.

4.1.1.4.2 Study 1. Objective 2: To apply this method to different sample sets, including both structured and unstructured quasispecies populations, and compare and complement the results with diversity estimates based on Hill numbers (qD).

Once the new method QFF was developed, we selected three data sets.

1.- A clone of the complete SARS-CoV2 spike gene, amplified using 12 overlapping amplicons, which is expected to show very low variability and therefore to be dominated by a highly prevalent master sequence surrounded with a minimal representation of mutant variants. As expected, the SARS-CoV-2 technical clone presented a dominant master haplotype with a median frequency of 0.876 (IQR 0.023). QFF (**Figure 9a**, corresponding to figure 1a in Article 1) showed no emergent haplotypes, whereas the low-fitness and very-low-fitness fractions (RHL) presented a median value of 0.11 (IQR 0.02) and 0.011 (IQR 0.0058), respectively. Hill numbers profile (**Figure 9b**, corresponding to Figure 1b in Article 1) showed a steep decrease from q=0 to q=1, remaining asymptotically flat from 1.5 onwards, with only a small difference from q=3 to q=∞.

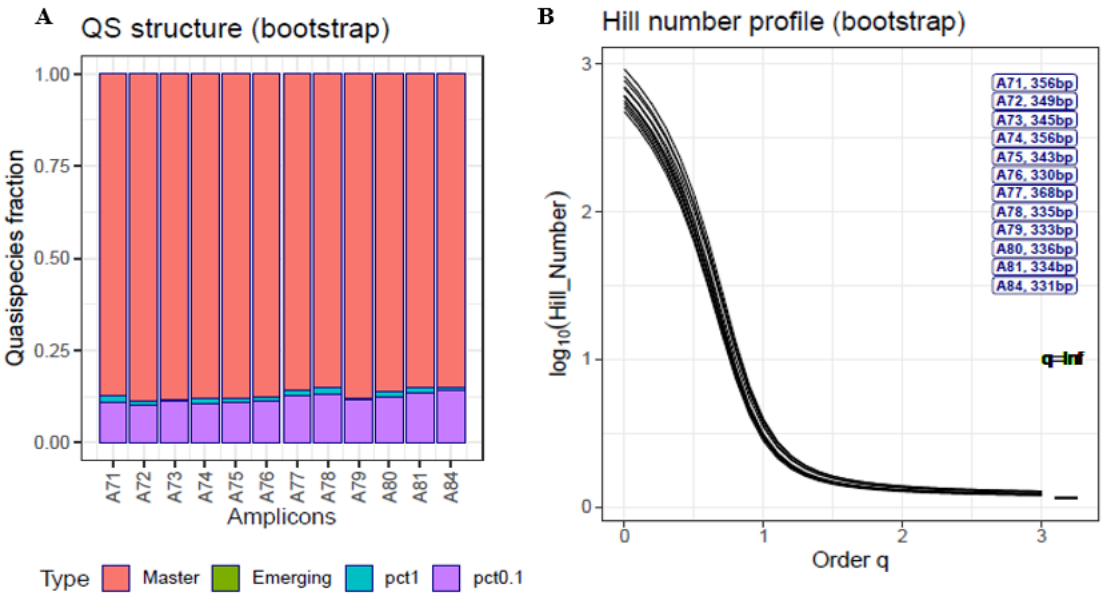


Figure 9. Quasispecies fitness fractions (9A) and Hill number profile (HNP) plots (9B) for a SARS-CoV-2 technical clone. The labels on the right of the HNP are sorted top-down in decreasing order of Hill numbers at q=2; the ARTIC amplicon nomenclature is used (A71 to A84). Pct1 low fitness (0.1<Freq<=1%), pct0.1 very low fitness (Freq<=0.1%). Each bar or curve corresponds to an amplicon of the S gene. Bootstrap values are represented. Coverage range: 51,102–299,349 reads; median 76,954 reads.

2.- A second data set, with expected moderate variability**,** comprising three amplicons of the NS5B region from a controlled experiment using cultured HCV-infected hepatoma cells (derived from an HCV plasmid chimera transcript) treated with mutagens or inhibitors. This study yielded richer plots than in the previous data set (**Figure 10**, corresponding to figure 2 in article 1), with different characteristics depending on the treatment. Inhibitor treatments resulted in lower fractions of emerging and RHL haplotypes compared with the control, as well as higher master haplotype frequencies. In contrast, mutagen exposure yielded to increased RHL fractions and a reduced master haplotype relative to the control. In **Figure 10b** (corresponding to figure 2b in Article 1), the HNP is shown with labels sorted top-down in decreasing order of Hill numbers at q=2. As the lower QFFs increase in volume and the master volume decreases, the curves remain at

higher levels, and show a larger difference between q=3 and q=∞. Some curves cross over each other.

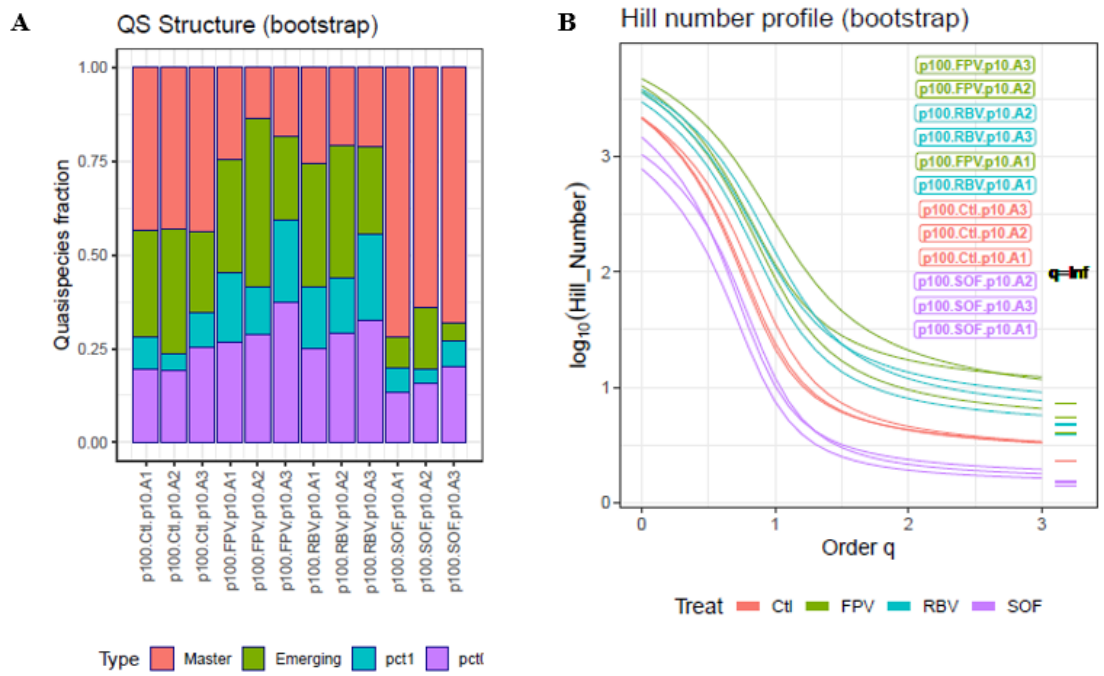


Figure 10. QFF (10a) and HNP (10b) plots for the HCV dataset. The labels on the right of the HNP are sorted top-down in decreasing order of Hill numbers at q=2. Ctl control; FPV favipiravir; RBV ribavirin; SOF sofosbuvir. A1 to A3 refer to the amplicons analysed. Pct1 low fitness (0.1<Freq<=1%), pct0.1 very low fitness (<=0.1%). Bootstrap values are represented. Coverage range 48,057-335,535 reads; median 206,354 reads.

3.- The third data set, characterized by extremely high variability**,** included the changes in the viral quasispecies observed following sequential RBV administration in a clinical case of HEV infection, as shown in **Figure 11a** (corresponding to figure 3a in Article 1). In the first sample (baseline), analysed on day 5 after the diagnosis (23 May 2018), the viral quasispecies already exhibited a high burden of rare haplotypes in both the <0.1% and 0.1%-1% frequency ranges, with the master haplotype accounting for less than 50%. Early RBV treatment (600mg) further increased the proportion of rare haplotypes and reduced the master sequence. After treatment interruption, the master haplotype temporarily predominated but at low viral load (3 logs). Subsequent viral rebound required the administration of higher

RBV dose (800mg), which reduced viral load but initially maintained quasispecies structure. During continued treatment, the master haplotype briefly recovered (>50%) as viral load decreased, followed by renewed gradual loss of the master sequence despite increasing the RBV dose to 1000 mg. Treatment was stopped as the master haplotype fell below 0.1. One year after the last end of treatment, the quasispecies was dominated by very low frequency haplotypes, with minimal master haplotype presence. This changes in diversity over time were reflected by shifts in Hill numbers (**Figure 11b**, corresponding to figure 3b in Article 1).

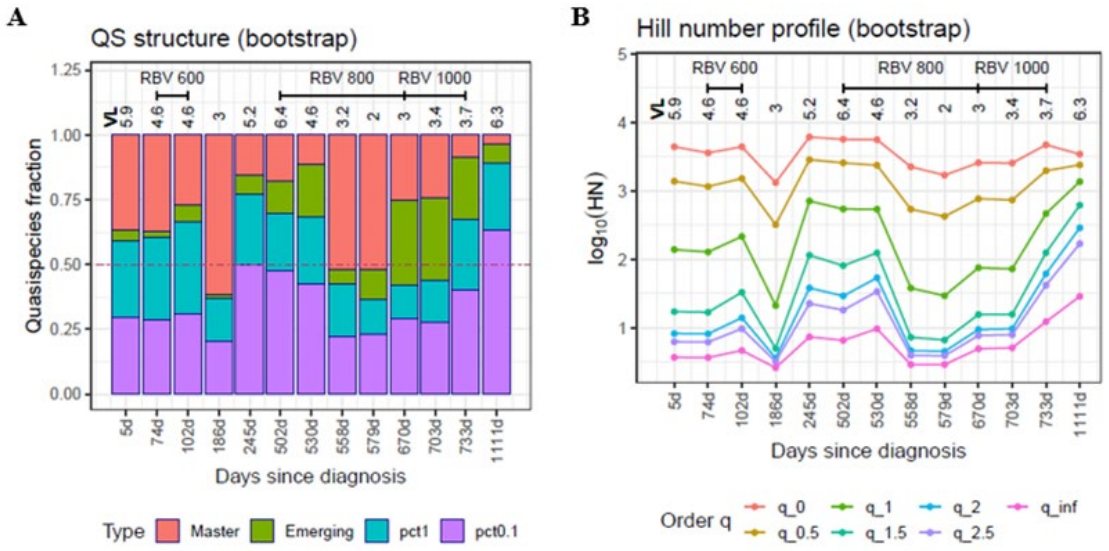


Figure 11. QFF (11a) and HNP (11b) plots for the HEV patient follow-up. The HNP is plotted here as cross sections of the profile at given q values. Each line corresponds to a q value; the lines show how this value changes over time. Bootstrap values are represented. On the x-axis, days since the diagnosis for each sample. Coverage range: 53,307-503,770 reads; median 328,271 reads.

This study showed that, at the end of the third RBV treatment, HEV exhibited an unstructured quasispecies at the nucleotide level, in which the master haplotype had a very low frequency and could be easily replaced by any mutant present in the emerging fraction (the fraction of molecules with a frequency higher than 1%). Interestingly, the phenotypic QFF analysis

revealed a completely different quasispecies structure. The last sample (1111 days post-diagnosis, 2 June 2021), showed a master phenotype accounting for >80.9% of the molecules. The master phenotype clearly predominated in the quasispecies throughout treatment, except in two samples (at end of treatment with RBV 800mg and during treatment with RBV 1000mg) where the predominant amino acid sequence changed (**Figure 12,** corresponding to Figure 5 in Article 1). At the other time points, the same master sequence predominated in the quasispecies. As these results were obtained from a single patient, they required further confirmation.

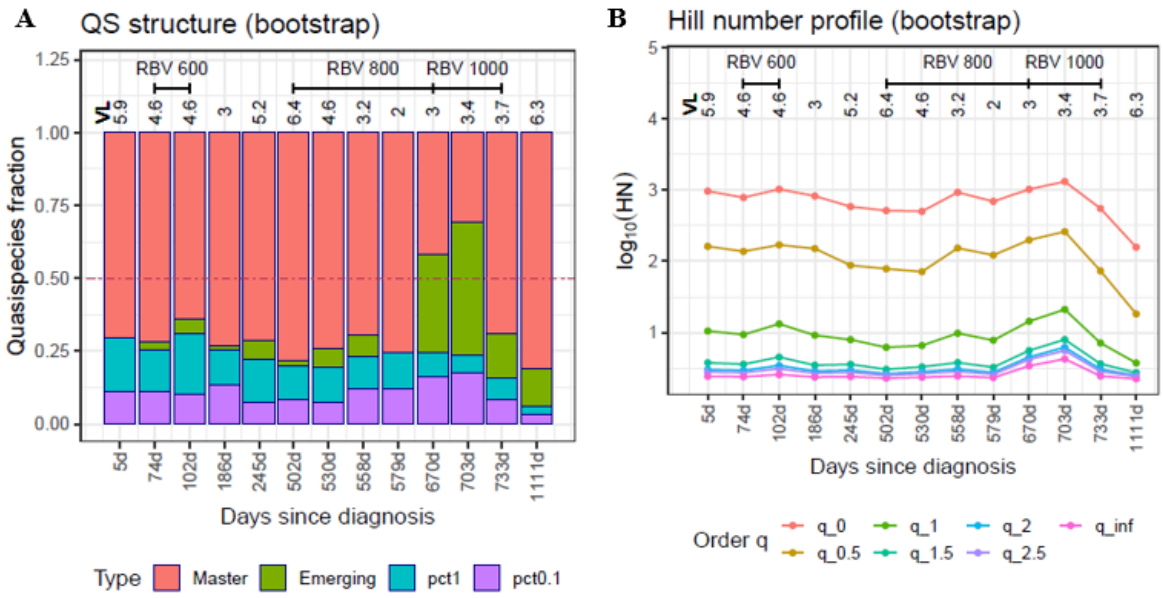


Figure 12. QFF (12a) and HNP (12b) plots for the quasispecies as amino acid haplotypes (phenotypes) for the HEV follow-up. Viral loads (VL) are expressed as logarithms. RBV, ribavirin. Pct1 low fitness (0.1<Freq<=1%), pct0.1 very low fitness (<=0.1%).

4.1.1.5 Study 1. Conclusions

- The combination of the Quasispecies Fitness Fraction Partition (QFF) and the Hill Number Profile (HNP) provides an easily interpretable visualization of quasispecies evolution, which may have significant clinical relevance.
- Quasispecies analysis using the novel QFF approach could serve as a useful tool to monitor viral populations ranging from very low to extremely high variability under mutagenic treatments during chronic disease progression.
- Most RBV-induced genetic diversity consists of synonymous mutations, as the master phenotype remains predominantly preserved across all longitudinal samples.

4.1.2 Article 1

Quasispecies Fitness Partition to Characterize the Molecular Status of a Viral Population. Negative Effect of Early Ribavirin Discontinuation in a Chronically Infected HEV Patient

Josep Gregori†, Sergi Colomer-Castell†, Carolina Campos, Marta Ibañez-Lligoña, Damir Garcia-Cehic, Ariadna Rando-Segura, Caroline Melanie Adombie, Rosa Pintó, Susanna Guix, Albert Bosch, Esteban Domingo, Isabel Gallego, Celia Perales, Maria Francesca Cortese, David Tabernero, Maria Buti, Mar Riveiro-Barciela, Juan Ignacio Esteban, Francisco Rodriguez-Frias and Josep Quer*

International Journal of
Molecular Sciences

MDPI

Article

Quasispecies Fitness Partition to Characterize the Molecular Status of a Viral Population. Negative Effect of Early Ribavirin Discontinuation in a Chronically Infected HEV Patient

Josep Gregori [1,2,*,†], Sergi Colomer-Castell [1,2,3,†], Carolina Campos [1,2,3], Marta Ibañez-Lligoña [1,2], Damir Garcia-Cehic [1,2], Ariadna Rando-Segura [2,4], Caroline Melanie Adombie [1,5], Rosa Pintó [6,7], Susanna Guix [6,7], Albert Bosch [6,7], Esteban Domingo [2,8], Isabel Gallego [8], Celia Perales [2,8,9], Maria Francesca Cortese [2,10], David Tabernero [2,10], Maria Buti [1,2,11], Mar Riveiro-Barciela [1,2,11], Juan Ignacio Esteban [1,2,11], Francisco Rodriguez-Frias [2,10] and Josep Quer [1,2,3,*]

1 Liver Diseases-Viral Hepatitis, Liver Unit, Vall d'Hebron Institut de Recerca (VHIR), Vall d'Hebron Barcelona Hospital Campus, Passeig Vall d'Hebron 119-129, 08035 Barcelona, Spain
2 Centro de Investigación Biomédica en Red de Enfermedades Hepáticas y Digestivas (CIBERehd), Instituto de Salud Carlos III, Av. Monforte de Lemos, 3-5, 28029 Madrid, Spain
3 Biochemistry and Molecular Biology Department, Universitat Autònoma de Barcelona (UAB), Campus de la UAB, Plaça Cívica, 08193 Cerdanyola del Vallès, Spain
4 Microbiology Department, Vall d'Hebron Institut de Recerca (VHIR), Vall d'Hebron Hospital Universitari, Vall d'Hebron Barcelona Hospital Campus, Passeig Vall d'Hebron 119-129, 08035 Barcelona, Spain
5 Institute of Agropastoral Management, University Peleforo Gon Coulibaly, Korhogo BP 1328, Côte d'Ivoire
6 Enteric Virus Laboratory, Section of Microbiology, Virology and Biotechnology, Department of Genetics, Microbiology and Statistics, School of Biology, University of Barcelona, 08028 Barcelona, Spain
7 Enteric Virus Laboratory, Institute of Nutrition and Food Safety (INSA), University of Barcelona, 08028 Barcelona, Spain
8 Centro de Biología Molecular "Severo Ochoa" (CBMSO, CSIC-UAM), Consejo Superior de Investigaciones Científicas (CSIC), Campus de Cantoblanco, 28049 Madrid, Spain
9 Department of Clinical Microbiology, Instituto de Investigación Sanitaria-Fundación Jiménez Díaz University Hospital, Universidad Autónoma de Madrid (IIS-FJD, UAM) Av. Reyes Católicos 2, 28040 Madrid, Spain
10 Biochemistry Department, Vall d'Hebron Institut de Recerca (VHIR), Vall d'Hebron Hospital Universitari, Vall d'Hebron Barcelona Hospital Campus, Passeig Vall d'Hebron 119-129, 08035 Barcelona, Spain
11 Medicine Department, Universitat Autònoma de Barcelona (UAB), Campus de la UAB, Plaça Cívica, 08193 Bellaterra, Spain
* Correspondence: josep.gregori@gmail.com (J.G.); josep.quer@vhir.org (J.Q.)
† These authors contributed equally to this work.

Abstract: The changes occurring in viral quasispecies populations during infection have been monitored using diversity indices, nucleotide diversity, and several other indices to summarize the quasispecies structure in a single value. In this study, we present a method to partition quasispecies haplotypes into four fractions according to their fitness: the master haplotype, rare haplotypes at two levels (those present at <0.1%, and those at 0.1–1%), and a fourth fraction that we term *emerging haplotypes*, present at frequencies >1%, but less than that of the master haplotype. We propose that by determining the changes occurring in the volume of the four quasispecies fitness fractions together with those of the Hill number profile we will be able to visualize and analyze the molecular changes in the composition of a quasispecies with time. To develop this concept, we used three data sets: a technical clone of the complete SARS-CoV-2 spike gene, a subset of data previously used in a study of rare haplotypes, and data from a clinical follow-up study of a patient chronically infected with HEV and treated with ribavirin. The viral response to ribavirin mutagenic treatment was selection of a rich set of synonymous haplotypes. The mutation spectrum was very complex at the nucleotide level, but at the protein (phenotypic/functional) level the pattern differed, showing a highly prevalent master phenotype. We discuss the putative implications of this observation in relation to mutagenic antiviral treatment.

Keywords: quasispecies; deep-sequencing; variability; rare haplotypes; fitness; mutagens

Citation: Gregori, J.; Colomer-Castell, S.; Campos, C.; Ibañez-Lligoña, M.; Garcia-Cehic, D.; Rando-Segura, A.; Adombie, C.M.; Pintó, R.; Guix, S.; Bosch, A.; et al. Quasispecies Fitness Partition to Characterize the Molecular Status of a Viral Population. Negative Effect of Early Ribavirin Discontinuation in a Chronically Infected HEV Patient. *Int. J. Mol. Sci.* **2022**, *23*, 14654. https://doi.org/10.3390/ijms232314654

Academic Editor: Rui Vitorino

Received: 5 October 2022
Accepted: 17 November 2022
Published: 24 November 2022

1. Introduction

Viral quasispecies [1] are intrinsically dynamic entities, with new genomes (haplotypes) being created during each replication cycle, mainly produced by the activity of an error-prone viral polymerase. The fate of each genome depends on its fitness [2]; that is, its capacity for replication in competition with other genomes in the quasispecies within the current environment. At a given time, the approximate fitness of a viral genome in a specific sample can be inferred from its frequency in the quasispecies [3,4]. That is, by the relative number of molecules belonging to the genome at that time point expressed as a fraction of the total. These frequencies can be considered a summary of the current molecular state of the quasispecies. The information gained by monitoring the molecular status of a viral quasispecies in an infection is particularly useful for following the viral response to a monoclonal antibody [5,6] or mutagenic agent [7,8]. Changes in the haplotype frequencies indicate the type of effects produced.

To characterize a viral quasispecies, we had to find an optimal balance between an analysis with high final coverage of shorter genomic fragments or with lower final coverage of larger fragments. The emergence of single-molecule real-time (SMRT) sequencing has opened the possibility to sequence large fragments, including entire viral genomes, in a single read as has been reported in influenza virus [9], hepatitis C virus [10], and human immunodeficiency virus [11]. However, SMRT technology typically has lower coverage (maximum 4M reads per run) and higher error rates than Illumina techniques [12,13]. SMRT is proven to have great value for assembling genomes [14] but requires further development to be useful for quasispecies analysis. For the present study, we needed very extensive final coverage ($\geq 10^5$ reads) to limit potential bias caused by a small sample size [3]. Hence, we used the MiSeq Illumina instrument, which enables analysis of amplicons in a size range of 300–500 base pairs (bp) at very high final depth and with acceptable error levels. The methods we propose are independent of the sequencing platform used. The only requirement is to have a set of high-quality haplotypes with their frequencies in the quasispecies. For this purpose, the sequencing data obtained here were treated to preserve the integrity of an amplicon's full-length reads; reads were either accepted or refused attending to their quality; however, they were never trimmed, except for the primers. Thus, the term *haplotype* used here refers to amplicon haplotypes, not to viral haplotypes, and the analyses are done amplicon-to-amplicon or on a single amplicon.

In a previous publication [7], we introduced the term *rare haplotype load* (RHL) in the context of a controlled experiment with HCV-infected hepatocytes treated with two mutagenic agents, the nucleoside analogues ribavirin (1-*β*-D-ribofuranosyl-1-*H*-1,2,4-triazole-3-carboxamide) and favipiravir (T-705; 6-fluoro-3-hydroxy-2-pirazinecarboxamide), and one inhibitor, sofosbuvir (isopropyl (2S)-2-[[[(2R,3R,4R,5R)-5-(2,4-dioxopyrimidin-1-yl)-4-fluoro-3-hydroxy-4-methyl-tetrahydrofuran-2-yl]methoxy-phenoxy-phosphoryl]amino]propanoate) [7]. The RHL refers to the fraction of genomic molecules in the quasispecies representing low- to very low-fitness haplotypes, and it was used as a biomarker of mutagenesis. In this study, we extend the concept of quasispecies partitioning according to their fitness (frequency within the population) into four fractions or haplotype categories: the master or dominant haplotype (present at the highest frequency within the quasispecies); the RHL divided into two levels, haplotypes present at less than 0.1% and haplotypes present from 0.1% to 1%; and a fourth category including emerging haplotypes, defined as those present at frequencies below the master value and above 1%. Emerging haplotypes are considered to have the potential to proliferate, attain higher frequencies, and possibly overtake the current master.

Here, we propose to represent quasispecies evolution as the changes observed in the volume (fraction of molecules) of the four quasispecies fitness fractions (QFF) combined with the Hill number profile (HNP) [15], which quantifies the effective number of haplotypes. This combination provides a means to visualize and analyze molecular changes in the composition of a quasispecies over time. Three data sets were used for this purpose: (1) a technical clone of the complete SARS-CoV-2 spike gene [16], sequenced using 12 overlapping amplicons; (2) a subset of data previously used in a study of the RHL as a marker of

lethal mutagenesis [7]; and (3) data from a clinical follow-up study of a patient chronically infected with hepatitis E virus (HEV) and treated with ribavirin (RBV), reported here for the first time.

2. Results

The technical clone is the simplest example (Figure 1). Despite its name, the commercial product carries multiple errors (although at low levels) due to the chemical yield at each synthesis step, which by definition never reaches 100% [17]. The errors observed include mainly deletions, which in the present case were bioinformatically corrected, but there were also some point mutations, possibly carry-overs from previous synthesis steps, which were left as they were. Some of these errors might also have been caused during the PCR amplification or sequencing steps. In Figure 1A, the QFF plot shows the <0.1% fraction which has a median of 0.11 (interquartile range [IQR] 0.02), the 0.1% to 1% fraction with a median of 0.011 (IQR 0.0058), no emerging haplotypes, and the master haplotype with a median of 0.876 (IQR 0.023). The HNP (Figure 1B) curve shows a steep decrease from $q = 0$ to $q = 1$. The curves remain asymptotically flat from $q = 1.5$ onwards, and there is only a small difference from $q = 3$ to $q = \infty$.

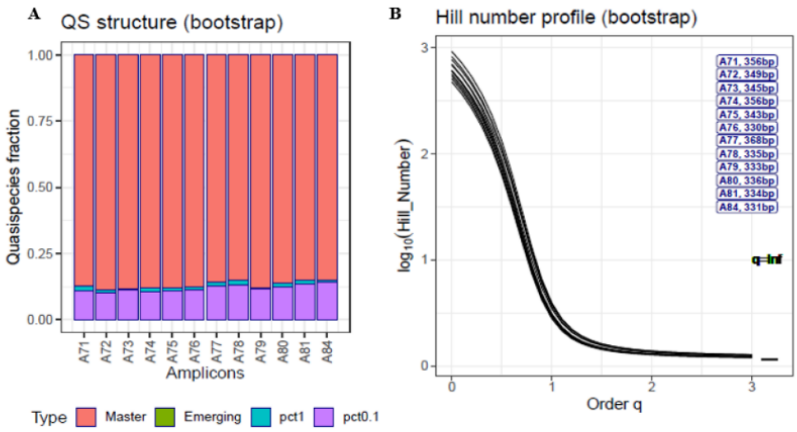


Figure 1. Quasispecies fitness fractions (**A**) and Hill number profile (HNP) plots (**B**) for a SARS-CoV-2 technical clone. The labels on the right of the HNP are sorted top-down in decreasing order of Hill number at $q = 2$; the ARTIC amplicon nomenclature is used (A71 to A84). Pct1 low fitness ($0.1 < \text{Freq} \leq 1\%$), pct0.1 very low fitness ($\text{Freq} \leq 0.1\%$). Each bar or curve corresponds to an amplicon of the S gene. Bootstrap values are represented. Coverage range: 51,102–299,349 reads; median 76,954 reads.

The controlled experiment with cultured HCV-infected human hepatoma cells treated with mutagens or inhibitors yielded richer plots (Figure 2A), which are characteristic of the effects of treatment. The median QFF and IQR values of the three amplicons for each fraction and under each condition are given in Table 1. Inhibitor treatment resulted in lower fractions of emerging haplotypes and RHL haplotypes with respect to the control, as well as higher master frequencies. In contrast, mutagen exposure yielded higher RHL fractions and a depressed master relative to the control. In Figure 2B, the HNP is shown with labels sorted top-down in decreasing order of Hill number at $q = 2$. As the lower QFFs increase in volume and the master volume decreases, the curves remain at higher levels, and show a larger difference between $q = 3$ and $q = \infty$. Some curves cross over each other. This generally happens when there is one quasispecies with a large number of haplotypes having limited diversity in the frequencies, and a second with a smaller number of haplotypes but with higher diversity in the frequencies.

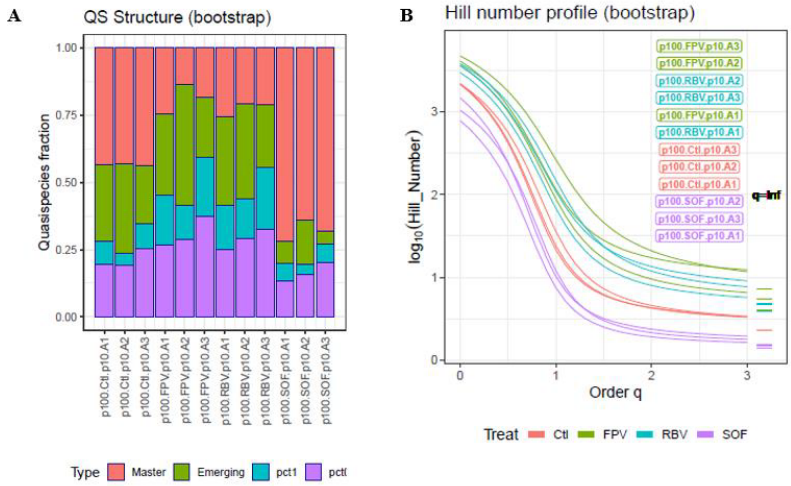


Figure 2. QFF (**A**) and HNP (B) plots for the HCV dataset. The labels on the right of the HNP are sorted top-down in decreasing order of Hill number at q = 2. Ctl control; FPV favipiravir; RBV ribavirin; SOF sofosbuvir. A1 to A3 refer to the amplicons analyzed. Pct1 low fitness (0.1 < Freq ≤ 1%), pct0.1 very low fitness (≤0.1%). Bootstrap values are represented. Coverage range 48,057–335,535 reads; median 206,354 reads.

Table 1. Median (Interquartile range) values of each fraction by treatment condition over the three amplicons.

	Master	Emerging	RHL_1_0.1	RHL_0.1
Control	0.436 (0.0030)	0.283 (0.0566)	0.085 (0.0232)	0.197 (0.0304)
FPV	0.184 (0.0541)	0.298 (0.111)	0.186 (0.0453)	0.290 (0.0519)
RBV	0.211 (0.0235)	0.331 (0.0586)	0.160 (0.0404)	0.293 (0.0366)
SOF	0.680 (0.0403)	0.081 (0.0577)	0.064 (0.0142)	0.158 (0.0348)

The changes occurring in the viral quasispecies following sequential RBV administration in the clinical case of HEV infection are shown in Figure 3A. In the first sample (baseline), analyzed on day 5 after the diagnosis (23 May 2018), the viral quasispecies already had a high burden of rare haplotypes at both <0.1% and 0.1–1%, with the master haplotype at <50%. The first mutagenic treatment (RBV 600mg) increased the RHL while reducing the master haplotype. Treatment was stopped on day 158 following the diagnosis, and 28 days later (20 November 2018), the master haplotype predominated in the quasispecies, but at a low viral load (3 logs).

Two months later (245 days since diagnosis, 18 January 2019), in the absence of treatment, the quasispecies had diversified to a higher level than was seen at baseline, with the 0.1% RHL reaching 50%. Nine months later (502 days since diagnosis, 2 October 2019), the viral load had increased to more than six logarithms, and a larger RBV dose (800mg) was prescribed. One month later (530 days since diagnosis, 30 October 2019), the quasispecies showed a structure very similar to that of the previous analysis in QFF terms, but the viral load had decreased by 2 logs. Suddenly, one month later (558 days since diagnosis, 27 November 2019) while the patient was still under treatment, the master sequence recovered to >50%, and remained at the same level for an additional 20 days (day 579 since diagnosis, 18 December 2019), whereas the viral load decreased to 2 logs. Three months later (670 days since diagnosis, 18 March 2020), the volume of emerging haplotypes had increased while the master haplotype showed a >50% decline with respect to the previous time point, with slightly higher viral loads. This structure was maintained for another month, at similar viral loads. Finally, one month later (733 days since diagnosis,

20 May 2020), when the master haplotype had further declined to <10%, treatment had to be stopped. One year later (1111 days since diagnosis, 2 June 2021), in the absence of treatment, the master haplotype was present at 3.6%, the emerging volume was 7.9%, and >88.5% of the quasispecies was comprised of haplotypes with frequencies < 1%. The HNP (Figure 3B) shows how the Hill numbers, at selected q values, changed over time.

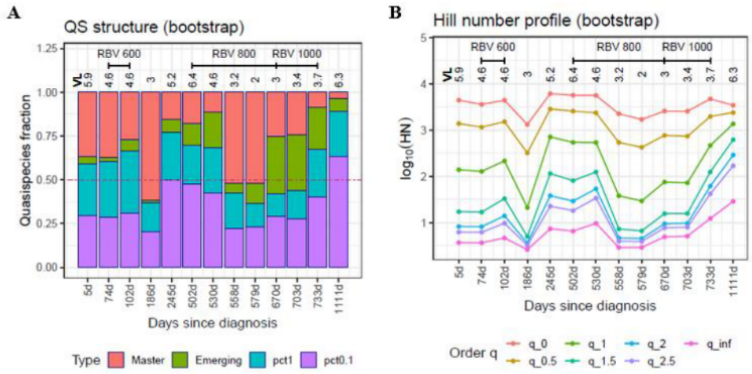


Figure 3. QFF (**A**) and HNP (**B**) plots for the HEV patient follow-up. The HNP is plotted here as cross-sections of the profile at given q values. Each line corresponds to a q value; the lines show how this value changes over time. Bootstrap values are represented. On the x-axis, days since the diagnosis for each sample. Coverage range: 53,307–503,770 reads; median 328,271 reads.

The UPGMA tree of the master sequences of all samples (Figure 4A) shows that the same master haplotype was maintained from the time of the diagnosis up to 530 days later (30 October 2019). From then on, the master differed at each time point except the last one (1111 days since diagnosis, 2 June 2021) when the master was the same as the sample at 670 days (18 March 2020). The UPGMA tree (Figure 4B) resulting from the Da population distances based on the top 50 haplotypes in each sample, displays a similar structure. The samples at 558 and 579 days since diagnosis (27 November 2019 and 18 December 2019) show a divergence in the structure and in the master sequence. These correspond to the lowest viral loads in the follow-up, with the master sequence predominating in the quasispecies.

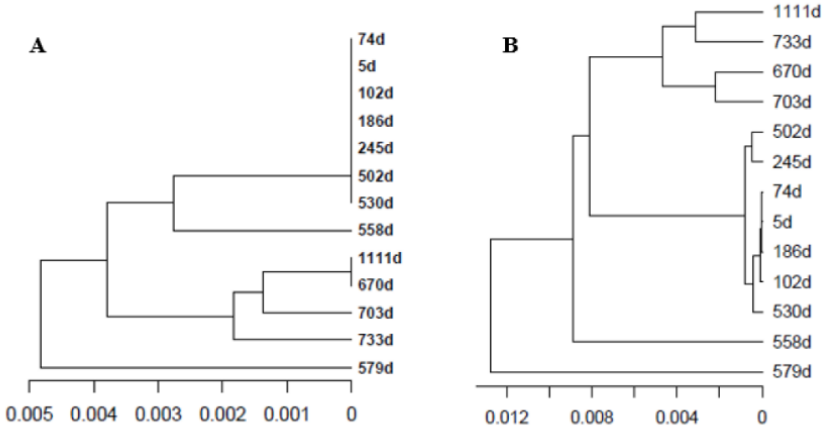


Figure 4. UPGMA tree of the master haplotypes based on raw nucleotide distances (**A**), and quasispecies tree based on the Da population distances (**B**) taking the top 50 haplotypes in each sample. Each sample is labeled as days since the diagnosis.

To understand why the profile of the last sample (1111 days since diagnosis, 2 June 2021), analyzed after one year with no treatment, showed a highly mutagenized quasispecies, the nucleotide haplotypes were translated to amino acids and recollapsed to obtain amino acid haplotypes (phenotypes) with their corresponding frequencies (Figure 5A,B). The last sample (1111 days since diagnosis, 2 June 2021) shows a master phenotype accounting for >80.9% of the molecules. The master phenotype clearly predominated in the quasispecies along treatment, except in two samples (at end of treatment with RBV 800 mg and during treatment with RBV 1000 mg), where there was a change in the predominant amino acid sequence (Figure S1a). At the other time points, the same master sequence predominated in the quasispecies. These findings indicate that although a quasispecies can have a highly mutated spectrum at the nucleotide level, it may remain almost unchanged at the amino acid level, suggesting that functionality is at least transiently maintained (Figures S1a,b and S2).

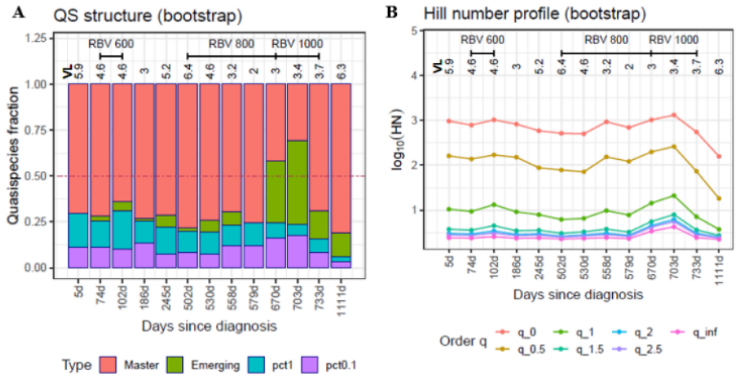


Figure 5. QFF (**A**) and HNP (**B**) plots for the quasispecies as amino acid haplotypes (phenotypes) for the HEV follow-up. Viral loads (VL) are expressed as logarithms. RBV, ribavirin. Pct1 low fitness ($0.1 < \text{Freq} \leq 1\%$), pct0.1 very low fitness ($\leq 0.1\%$).

The number of synonymous haplotypes corresponding to the master phenotype steeply increased after the treatment discontinuations (Figure 6). That is, the various haplotypes identical to the master phenotype generated during treatment were able to easily increase in frequency when it was stopped, as they all had highest functional fitness.

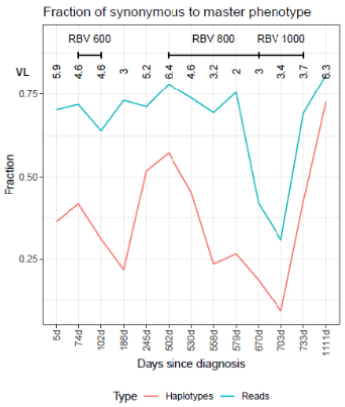


Figure 6. Fraction of haplotypes synonymous to the master phenotype (orange), and fraction of reads for these haplotypes (green). On the x-axis, days since the diagnosis for each sample. Viral loads (VL) are expressed as logarithms. RBV, ribavirin.

A UPGMA tree of master phenotypes based on their Grantham amino acid distances, and a quasispecies tree based on Da population distances were constructed using the top 20 most frequent phenotypes in each sample (Supplementary Figures S1a,b and S2). Apparently, viral treatment led to production of a rich set of haplotypes synonymous to the master phenotype, all with equal or comparable fitness, which proliferated when treatment was discontinued. This was evidenced by the decline in the master haplotype volume and the increase in emerging haplotypes synonymous to the master. As a whole, the resulting quasispecies might be more resistant to further mutagenic treatment and better fit to its current environment. A large number of highly fit haplotypes could correspond to a large number of molecular pathways to escape a treatment.

A final illustration of this conclusion is provided in Supplementary Figures S3 and S4, where Montserrat plots depict the distribution of the 1000 most abundant haplotypes and all phenotypes in the last sample (1111 days since diagnosis, 2 June 2020). In Montserrat plots, haplotypes are sorted by number of substitutions with respect to the master first, and by decreasing order of abundance second [18]. Each successive peak represents additional differences with respect to the master. In correspondence, Supplementary Figure S5 shows the mean number of differences with respect to the master haplotype/phenotype per read, at both the nucleotide level (substitution load) and amino acid level (mutation load). Notably, the ratio of the nucleotide substitution load to the amino acid mutation load was 13.31 in the last sample (one year without treatment), only 1.24 in the baseline sample, shortly after diagnosis, and 6.57 when the last treatment was discontinued. In addition, the mutation load (amino acid level) in the last sample was the lowest value in the series, despite the highly diverse quasispecies.

3. Discussion

The presence of a broad repertoire of genomes in the mutant spectra of RNA viruses represents a challenge for annotating and describing the genomes and the neighbor relationships among them. Several procedures have been developed to rank subgroups of related sequences within mutant spectra [19–23]. Alternative approaches have monitored mutant spectrum composition and diversification through quantification of diversity indices [4,24] and haplotype mapping using two-dimensional neural networks [25,26].

In our previous studies we introduced a number of new diversity indices, which, when adequately combined, provided information on both the abundance and divergence of haplotypes within a viral population [4]. In the present report, we have gone one step further, creating a procedure to divide a quasispecies population into four fractions using a fitness partition (QFF). The advantages of this approach include better statistical properties than most diversity indices and inclusion of the biological features of the population. The QFF procedure goes beyond a previous proposal restricted to the two fractions at lowest frequency in the mutant ensemble [7]. In addition, we used the Hill number profile [15] to provide a complementary view of the quasispecies structure. The profile is an enriched summary that assigns increasing weights to the haplotype frequencies in the quasispecies as the order q increases. In addition, the units provided (number of equally fit haplotypes) contribute to the interpretation of the results.

Bootstrapping provided an efficient down-sampling of mutant distributions in the quasispecies samples to the minimum coverage, thereby allowing comparisons regardless of the sample size. Hill numbers of orders below two are highly dependent on sample size and must be corrected. However, it is advisable to correct all diversity values. QFF is compatible with other ranking procedures used to study quasispecies, as the proposed fractioning can be applied to genome subsets obtained by other means. In particular, it can be a useful complement to self-organized fitness maps based on artificial neural networks [26]. It may also provide support to precisely define the SARS-CoV-2 mutant spectra which, according to recent results, are populated by a large proportion of low-frequency haplotypes [27,28].

The rationale behind the QFF definition and its main contribution to quasispecies analysis resides in the biological meaning of each fraction. Observation of a significant fraction of molecules corresponding to very low fitness haplotypes is only consistent with the presence of a mechanism that can generate mutants at a high rate, but cannot increase their frequency relative to competing genomes. Furthermore, concerning the response of HEV to RBV, the course of any treatment giving rise to resistant variants is only consistent with a decline in volume of the master haplotype together with a parallel increase in emerging haplotypes.

We used both the QFF and HNP to visualize and analyze two simple case models: the first, a SARS-CoV-2 technical clone, and the second, populations from a controlled experiment in which a clonal HCV population was serially passaged in cultured human hepatoma Huh-7.5 cells. These two datasets were used to demonstrate the capability of the proposed QFF/HNP combined method. Finally, the procedure was applied to a complex clinical case: follow-up of an HEV-infected patient treated with various RBV doses with two discontinuations, to illustrate its performance for clinical purposes. As this is a single example, the discussion provided for this case should be considered explanatory of how the information obtained with the QFF procedure can be of clinical value.

Lethal mutagenesis is a useful antiviral approach that consists of driving viral genomes to extinction—pushing the virus to cross the error catastrophe threshold by increasing the viral mutation rate above the maximum level compatible with infectivity—without mutagenizing the host cells [29–31]. A recent example of successful application of lethal mutagenesis is the use of molnupiravir against COVID-19 [32]. Several other cases have been reviewed [8]. Currently, no specific drugs have been approved for HEV infection; RBV is the main option as an off-label drug. RBV is a broad-spectrum mutagenic agent [33–35] that can increase mutation rates and result in extinction of the virus by lethal mutagenesis [36]. However, during RBV treatment, a reduction in the effective antiviral dose by low adherence or early treatment interruption can allow residual viral replication and production of rescue variants with decreased RBV sensitivity or altered replication fitness [37–39]. This could lead to selection of mutations resistant to RBV and the appearance of resistant variants.

In our study, deep-sequencing of samples from an HEV-infected patient receiving RBV treatment showed that the mutagenic agent led to a reduction in the most highly represented sequence (master) at the nucleotide level, together with a significant increase in the number of rare haplotypes, findings in agreement with a mutagenic effect of RBV on this virus. When the first round of RBV treatment (600 mg) was stopped, HEV relapsed, showing an increase in viral load and re-acquisition of the master sequence that had predominated before treatment was started. In the second and third rounds of RBV treatment, with increases in the drug concentration to 800 mg and 1000 mg and evidence of good adherence to therapy, the viral load remained unchanged and even increased, while the master sequence decreased once again. Notably, after stopping 1000 mg RBV treatment, the frequency of the master genome declined even further despite a three-log viral load increase.

However, when we examined the quasispecies at the protein (phenotypic/functional) level, the pattern drastically changed, showing a highly predominant master phenotype, despite the complexity of the mutation spectrum observed at the nucleotide level. On stopping treatment, the synonymous variants generated had the same replication capacity as the original master genome and were able to proliferate, leading to a high diversity of genomes that could all express the same phenotype. The reason why the dynamics of the mutant spectrum involved haplotypes with silent mutations is unknown. Furthermore, we cannot exclude that some of the variants produced might have had lower sensitivity to RBV. Both these reasons; a large reservoir of functional genomes and decreased sensitivity to treatment could explain the viral load increase in the presence of a high dose of mutagen. Failures in HEV RBV treatment have been described, but detection of RBV-resistant mutations requires sequencing the full HEV genome [38–40]. In our case, we assumed that the

effects of RBV on the sequenced amplicon would be similar to the impact on the remainder of the genome, as the virus cannot drive the mutagenic effects. Future approaches using the new SMRT circular consensus sequencing technology may provide further support of this amplicon to genome relationship.

The mutagenic effect of RBV on HEV could lead to viral extinction, but it also involves a risk of accumulating advantageous mutations and selecting fitness-enhancing ones [36]. In our study, while the patient was receiving antiviral treatment, a number of synonymous haplotypes were produced, which as a whole seemed to be stronger against further treatment and improved accommodation of the quasispecies to its current environment. The findings from our patient suggest that mutagenic antiviral therapy should ideally be combined with other antivirals. When this is not possible, as in HEV infection, treatment should be maintained, even after serum RNA tests negative, to avoid relapses which could lead to selection of fitness-enhancing mutations and treatment failure.

The QFF approach, although simple and straightforward, has some technical limitations, mainly due to the current state of high-throughput sequencing technology. The sequence length and error level are two aspects that limit each other. We can sequence amplicons up to slightly more than 500 bp in length with an acceptable error level using paired-end technology, but we cannot sequence full-length viral genomes of a few thousand base pairs at high depth (10^4–10^5) with low error levels and high coverage. In this study, we used amplicons larger than 300 bp and assumed that either the amplicon underwent effects similar to those that would occur in the whole genome, or that the amplicon was the target to study. Another limitation relates to a factor observed in all *-omics*, where the experimental design is of the utmost importance to avoid bias. Even in balanced designs, batch effects should be taken into account. In our case, we were not interested in a detailed account of point mutations and indels; rather, we aimed to provide a picture of the macroscopic structure of a quasispecies. This information is of value, as high-resolution deep sequencing unveils myriads of low frequency mutations that should correspond to an extensive repertoire of minority genomes [28]. In providing this general view, we can accept a certain presence of artifactual haplotypes, provided that all samples in the experiment show the same noise level, hence the need for an accurate experimental design. Filtering above the error level to avoid all artifacts would involve a considerable loss of information and jeopardize the type of analysis we are proposing. The impact of filtering all haplotypes below 0.1% or 1% on the total reads number, that is the information loss, can be seen in Figures 1–3. Nevertheless, low-level filters of very few reads per haplotype (e.g., 1–10 reads at 1×10^5 coverage) will have an impact on the number of haplotypes—that is, on the Hill numbers of low order (<1.5)—but will have a minimum effect on the number of reads. It could be helpful to perform a sensitivity analysis of the various diversity indices used with respect to this threshold. These limitations and warnings are equally applicable to any quasispecies study, whatever the indices or variables used, and are not exclusive to the methods proposed here.

We propose a simple method for monitoring the changes occurring in a quasispecies at the molecular level, involving fitness fractions and the Hill number profile. This combined method, which provides an easily interpretable visualization of quasispecies evolution in viral terms, was applied to two simple cases as a demonstration, and to samples from an RBV-treated HEV patient to illustrate its clinical value. The method is based on bioinformatic treatment of sequencing data to obtain a set of high-quality amplicon haplotypes with their corresponding frequencies to represent the quasispecies structure. Use of next-generation sequencing technology in combination with a good experimental design provides exceptional opportunities to study complex quasispecies and follow their evolution at the molecular level, in both research laboratories and clinical settings.

4. Materials and Methods

4.1. Samples

Two datasets were used to develop quasispecies molecular characterization with QFFs and the HNP:

- A technical clone of the SARS-CoV-2 *S* gene (Twist Synthetic SARS-CoV-2 RNA Control 2 MN908947.3, TWIST Biosciences, South San Francisco, CA, USA) sequenced in 12 amplicons [17]. Commercial Twist Synthetic SARS-CoV-2 RNA controls consist of six non-overlapping 5-Kb fragments generated from in vitro transcription of gene fragments. The synthetic controls were diluted at 1:10 to a concentration of 1×10^5 copies per microliter, PCR-amplified following the Sub-ARTIC v3 protocol [41] using a set of 28 primers (A71 to A84) covering the full *S* gene, and sequenced on a MiSeq Illumina system [42]. The haplotypes and corresponding frequencies in this analysis included all haplotypes common to both DNA strands after a previous filter at 2 reads. That is, at a minimum of 2 + 2 reads.
- Three HCV amplicons from samples taken from a controlled experiment, in which HCV-infected human hepatoma cells were observed in the presence or absence of RBV, favipiravir, or sofosbuvir [43,44]. Briefly, HCV p0 was the parental viral population obtained by electroporation of a transcript of plasmid Jc1FLAG2(p7-nsGluc2A) (a chimera of J6 and JFH-1, genotype 2a) [45] into Huh-7.5-Lunet cells and amplification in Huh-7.5 cells [46]. HCV p100 resulted from passaging the HCV p0 population 100 times in Huh-7.5 reporter cells [47]. HCV p100 was subsequently passaged 10 additional times in the presence of favipiravir (T-705) (Atomax Chemicals Co., Ltd., Shenzhen, China), RBV (Sigma, Kawasaki, Japan), or sofosbuvir. Drug concentrations were adjusted to produce comparable inhibition of HCV p0 progeny production. The amplicons sequenced covered the following HCV genomic regions: A1, spanning genomic residues 7626 to 7962; A2, residues 7941 to 8257; and A3, residues 8229 to 8653. The haplotypes and corresponding frequencies in this analysis included all haplotypes common to both strands, with no previous abundance filter; that is, a minimum of 1 + 1 reads.

In addition, we describe the quasispecies findings from the clinical follow-up case of a 27-year-old patient who acquired chronic HEV infection after undergoing two kidney transplantations. The patient received three different RBV regimens (Table 2). First, 600 mg per day for 3 months, which led to a significant reduction in viral load without achieving undetectable HEV RNA. Treatment was stopped. The patient relapsed, and a second treatment with RBV 800 mg daily was prescribed, with a new reduction in HEV levels. At month 5, RBV dosage was increased to 1000 mg daily for two additional months. Treatment was discontinued because of a lack of antiviral response, and viral load jumped three logs at 10 days after stopping treatment. A single amplicon covering genomic positions 6323 to 6734 on the ORF2 region was sequenced. The haplotypes and corresponding frequencies in this analysis included all haplotypes common to both DNA strands after a previous filter at 2 reads. That is, a minimum of 2 + 2 reads.

Table 2. Follow-up data, with dates and intervals, viral loads expressed as logarithms, and clinical observations. EOT, end of treatment.

Date (Y-M-D)	Interval (Days)	Days Since Diagnosis	Sample ID	LogVL	Observations
2018-05-18	0	0		5.91	Diagnosis
2018-05-23	5	5	S01	5.87	
2018-07-31	69	74	S03	4.60	Ribavirin 600 mg
2018-08-28	28	102	S04	4.60	
2018-10-23	56	158		1.54	EOT

Table 2. *Cont.*

Date (Y-M-D)	Interval (Days)	Days Since Diagnosis	Sample ID	LogVL	Observations
2018-11-20	28	186	S06	3.04	Relapse
2019-01-18	59	245	S08	5.18	
2019-10-02	257	502	S10	6.43	Ribavirin 800 mg
2019-10-30	28	530	S12	4.62	
2019-11-27	28	558	S14	3.18	
2019-12-18	21	579	S16	2.04	
2020-03-18	91	670	S17	3.04	Ribavirin 1000 mg
2020-04-20	33	703	S18	3.40	
2020-05-20	30	733	S20	3.68	EOT
2020-06-17	28	761		4.45	Relapse
2021-06-02	350	1111	S24	6.28	

4.2. Processing the Sequencing Data

The aim of the sequencing data treatment was to discard error-bearing reads while preserving full-length read integrity, so that haplotypes that completely cover the amplicon with their respective frequencies were incorporated. The steps in this process are the following:

- Obtain Fastq files with Illumina 2 × 300-bp paired-end reads;
- Recover full amplicon reads with FLASH [48] (min. 20-bp overlap, max. 10% mismatches). The 300-bp reads, when overlapped, result in reads covering complete ~400–500 bp amplicons;
- Remove full reads with 5% or more bases below a Phred score of Q30;
- Demultiplex and trim primers (max three differences accepted);
- Collapse reads (molecules) to haplotypes (amplicon-genomes) and their frequencies. The frequencies were calculated per haplotype of each amplicon;
- In certain cases, remove all haplotypes below a fixed frequency threshold;
- Remove all haplotypes that are not common to both DNA strands.

The final obtained haplotypes and their frequencies were the basis for all further calculations.

The SARS-CoV2 dataset consists of 12 amplicons (min 330 bp, max 368 bp, median 340 bp). The HCV dataset consists of three amplicons (312, 318, and 423 bp). Finally, the HEV study is based on single 363-bp amplicons (Supplementary Table S1). The amplicon sizes provided are the final result after primer trimming.

4.3. Quasispecies Fitness Partitions

At a given time, a quasispecies is usually comprised of a highly predominant haplotype, a few low- to medium-frequency genomes, various rare haplotypes with very low fitness but still able to replicate to some level, and some defective genomes unable to replicate. This composition can be modeled using the set of frequencies of all haplotypes as parameters of a multinomial distribution (Equation (1)),

$$\Pi = \{p_1, p_2, \ldots, p_n\} with \sum_{i}^{n} p_i = 1 \quad (1)$$

where p_1, p_2, … , p_n represent the various haplotypes, arranged in order of decreasing frequency. The parameters, p_i, are sorted in decreasing order without a loss of generality. In

this way, the quasispecies can be partitioned into fractions limited by frequency thresholds of interest, as in Equation (2), where a partition into four fractions is illustrated,

$$\begin{gathered} \Pi_1 = \{p_1, p_2, \ldots, p_k\}, \forall p_i : p_i \geq p_k \\ \Pi_2 = \{p_{k+1}, p_{k+2}, \ldots, p_l\}, \forall p_i : p_l \leq p_i < p_k \\ \Pi_3 = \{p_{l+1}, p_{l+2}, \ldots, p_m\}, \forall p_i : p_m \leq p_i < p_l \\ \Pi_4 = \{p_{m+1}, p_{m+2}, \ldots, p_n\}, \forall p_i : p_n \leq p_i < p_m \\ p'_1 = \sum_{i}^{k} p_i; p'_2 = \sum_{k+1}^{l} p_i; p'_3 = \sum_{l+1}^{m} p_i; p'_4 = \sum_{m+1}^{n} p_i; with \sum_{i=1}^{4} p'_i = 1 \end{gathered} \tag{2}$$

and where, p'_1, p'_2, p'_3, and p'_4 represent the four fractions.

In the typical quasispecies structure mentioned above, the four fractions can be defined as follows:

- **Master**: the fraction of molecules belonging to the most frequent haplotype; that is, the one present at the highest percentage ($p'_1 = p_1$);
- **Emerging**: the fraction of molecules present at a frequency > 0.1% and less than the master percentage, belonging to haplotypes that are able to compete with the predominant one and possibly replace it (p'_2);
- **Low fitness**: the fraction of molecules present at frequencies from 1% to 0.1%, belonging to haplotypes that have a low probability of progressing to higher frequencies (p'_3);
- **Very low fitness**: the fraction of molecules present at frequencies < 0.1% belonging to haplotypes with very low fitness and to defective genomes. The likely fate of these molecules individually is degradation, but the fraction is continuously fed with new very low fitness genomes produced by replication errors or by host editing activities (p'_4).

The evolutionary trends occurring in a viral quasispecies can be characterized by determining the changes that take place in the molecular volume of these fractions as a function of time.

The coefficient of variation (CV) of a proportion, p, for a given sample size, N (i.e., the standard deviation expressed in expected value units), is given in Equation (3). When the proportion is very small, it can be approximated by calculating the square root of the inverse of the product of the sample size multiplied by the proportion, as in Equation (4),

$$CV\left[\frac{x}{N}\right] = CV(p, N) = \frac{sd(p, N)}{p} = \sqrt{\frac{(1-p)}{Np}} \tag{3}$$

$$CV(p, N) \approx \sqrt{\frac{1}{Np}} \tag{4}$$

where x is the observed count, N is the sample size, and p is the observed proportion. In the experiments described in this study, the coverage (sample size, N) ranged from 10^4 to 10^6, with the average larger than 1×10^5, and our aim was to observe haplotypes present in very low proportions (p); that is, <0.1% (<$\times 10^{-3}$). Individually, haplotypes considered to have very low fitness will show a high CV, which means that some of them can be easily overlooked in a single sample. Nevertheless, when grouped together, as is seen in the above partition (the p' in Equation (2)), they amount to a much higher proportion than when counted individually and become more statistically stable. Thus, the fraction of molecules in the quasispecies belonging to haplotypes having very low fitness, p'_4, becomes more stable to sampling and less dependent on the sample size [7] than any single haplotype at these frequencies.

4.4. Hill Numbers

In addition to the partition described above, we can determine the diversity profile of a quasispecies with the use of Hill numbers [4,15,49]. Based on the expression of the generalized diversity index, of order q, ${}^qH(p)$ is given in Equation (5),

$$ {}^qH(p) = \sum_{i=1}^{H} p_i^q \tag{5}$$

the Hill number of order q, ${}^qD$, of a quasispecies corresponds to the number of equally fit haplotypes comprising a quasispecies with the same general diversity, ${}^qH$, as the original quasispecies, as is shown in Equation (6).

$$ \sum_{i=1}^{H} p_i^q = \sum^{D} \left(\frac{1}{D}\right)^q = D\left(\frac{1}{D}\right)^q = D^{1-q} \tag{6}$$

This results in Equation (7),

$$ {}^qD(p) = \left(\sum_{i=1}^{H} p_i^q\right)^{1/(1-q)} \tag{7}$$

where ${}^qD(p)$ is the Hill number of order q, calculated from the haplotype frequencies observed in the quasispecies, p_i. The diversity indices, ${}^qD(p)$, obey the replication principle [4] and are expressed in intuitive units (number of equally fit haplotypes). In ecology, the replication principle states that if we have N equally large, equally diverse groups (quasispecies), and no species (genomes) in common, the diversity of the pooled groups must be N times the diversity of a single group.

At increasing values of q starting from 0, we obtain diversity values that are also a transformation of other classical diversity indices:

- At $q = 0$, the Hill number is the number of haplotypes;
- at $q = 1$, it corresponds to the exponential of Shannon entropy;
- at $q = 2$, it is the inverse of the Simpson index; and
- at $q = \infty$, it is the inverse of the predominant haplotype.

The Hill number profile of a quasispecies is the curve we obtain from $q = 0$ to 3, plus the value at infinity. The curve becomes asymptotic beyond order 2 and reaches its minimum at infinity. As a result of the quasispecies values spanning a large range—more than three orders of magnitude—the Hill number profile is best represented in log_{10} units.

4.5. Abundance Filter Effect on Haplotype Distribution

The goal of step 6 in the sequencing data treatment described above is usually to limit technical errors (PCR + sequencing) to a level suitable for the purposes of the study, while maintaining the integrity of full amplicon reads. The required frequency threshold can be established through the use of clones that have been processed in parallel with the clinical samples and sequenced in the same run [50].

The immediate effect of this filter is removal of all haplotypes with abundances below the threshold, which in probabilistic terms, corresponds to truncating the distribution. Let p_k be this threshold, so that the haplotypes removed are those at frequencies p_{k+1}, p_{k+2}, $\ldots$, p_n. The resulting truncated quasispecies will show haplotype frequencies resulting from normalization of the remaining haplotype frequencies, $(p_1, p_2, \ldots, p_k)$, as described in Equation (8),

$$ \Pi' = (p_1, p_2, \ldots, p_k) / \sum_{i=1}^{k} p_i \tag{8}$$

where p_i are the frequencies calculated from read counts before the filter. In this manner, the original frequencies are simply re-scaled. The truncation represents a loss of information,

in the sense that below the error level, whatever it may be, there are authentic haplotypes that would equally be rejected.

4.6. Sample Size Dependence

Diversity indices are dependent to a varying extent on the sample size [4], and this dependence has to be taken into account when comparing values from different samples. In the present study, we used bootstrap to correct differences due to sample size. From the set of samples to be compared, the minimum coverage was taken as the sample size reference. Each sample then underwent 1000 bootstrap cycles (i.e., sampling with repositioning), taking the frequencies of all haplotypes in the sample as the probabilities, and the minimum coverage in the set of samples to be compared as the sample size. In each cycle, each diversity index was calculated from the resulting bootstrap sampling. At the end of the 1000 cycles, averages and standard deviations were computed for each diversity index in the study. In this way all diversity index values were referenced to the same sample size (coverage).

In previous research, we suggested a faster alternative to this process, calling it down-sampling with fringe trimming [3,4]. Although it provided good correction of sample size bias, it resulted in a loss of information that can be critical in some situations.

4.7. Distance between Quasispecies, Quasispecies Dendrograms, and Multidimensional Scale Plots (MDS)

A quasispecies can be seen as a genetic population, and the methods used to study diversification in a genetic population can also be used to determine the distance or dissimilarity between different quasispecies. There are several useful methods to quantify distance. When examining the Hamming distance, or any genetic distance, between pairs of haplotypes in two populations (quasispecies), the nucleotide divergence (Da) formula by Matoshi Nei [51] measures the net genetic distance between them, correcting the full genetic distance by subtracting their mean intra-population genetic diversity. This same method can be applied to calculate the inter-quasispecies phenotype distance, in this case substituting the matrices of genetic distances between pairs of haplotypes for the matrices of distances between amino acid haplotypes. The distance between proteins can be computed by the method of Grishin [52] to obtain dissimilarities between proteins from substitution matrices, such as PAM or BLOSUM. Alternatively they can be computed directly from matrices of distances between pairs of amino acids, using methods such as that of Fitch [53], or Grantham [54]. When the quasispecies are closely related, as in the HEV follow-up study here, an alternative type of distance or dissimilarity of interest can be obtained directly from the haplotype frequencies of the two quasispecies, regardless of the genetic distance between them, by applying the method of Yue—Clayton [55]. This method can also be applied to the quasispecies fitness fractions introduced above.

The distances or dissimilarities obtained can be used to plot quasispecies dendrograms or trees, or multidimensional scaling maps, to help visualize how the quasispecies in a study are related.

4.8. Software and Statistics

All computations were done in R (v4.0.3) [56] with in-house scripts, using the Biostrings [57], ShortRead [58], and QSutils [59] packages from Bioconductor [60], as well as ape [61], tidyverse [62], and ggplot2 [63].

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/ijms232314654/s1.

Author Contributions: J.G.: Conceptualization, methodology, software, formal data analysis, visualization, writing original draft; S.C.-C.: Methodology, investigation, visualization, writing original draft; C.C.: Investigation, visualization; M.I.-L.: Software, data curation; D.G.-C.: Methodology, investigation, project administration; A.R.-S.: Samples, investigation, and resources; C.M.A.: Investigation,

visualization; R.P.: Conceptualization, resources, writing review and editing; S.G.: Methodology, writing—review; A.B.: Supervision, funding acquisition; E.D.: Supervision, writing—review and editing; I.G.: Resources, data curation; C.P.: Resources, formal analysis; M.F.C.: Methodology, visualization; D.T.: Investigation, writing editing; M.B.: Conceptualization, resources, writing—review and editing; M.R.-B.: Resources, data analysis, validation; J.I.E.: Conceptualization, resources, validation, writing review and editing; F.R.-F.: Conceptualization, funding acquisition, data analysis, writing review and editing; J.Q.: Conceptualization, methodology, validation, supervision, funding acquisition, and writing original draft. All authors have read and agreed to the published version of the manuscript.

Funding: This study was partially supported by Plan Estratègic de Recerca i Innovació en Salut (PERIS)—Direcció General de Recerca i Innovació en Salut (DGRIS), Catalan Health Ministry, Generalitat de Catalunya; Centro para el Desarrollo Tecnológico Industrial (CDTI) from the Spanish Ministry of Economy and Business, grant number IDI-20200297; grant PI19/00301 from Instituto de Salud Carlos III cofinanced by the European Regional Development Fund (ERDF), and Gilead's biomedical research project GLD21/00006. Work at CBMSO in Madrid was supported by grant PID2020-113888RB-100 from Ministerio de Ciencia e Innovación and S2018/BAA-4370 (PLATESA2 from Comunidad de Madrid/FEDER).

Institutional Review Board Statement: The study was approved by Vall d'Hebron University Hospital Ethics Committee, with reference number PR(AG)259-2020.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The genomic nucleotide sequences included in this study have been submitted in the GENBank repository database as BioProject ID PRJNA876218.

Acknowledgments: We thank Celine Cavallo for English language support.

Conflicts of Interest: The authors declare no conflict of interest.

References

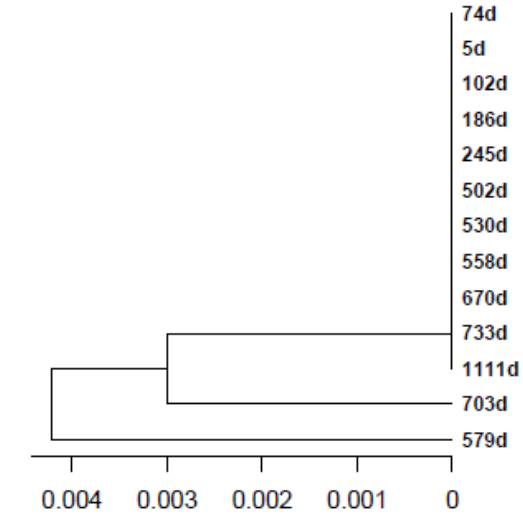
1. Domingo, E.; Escarmis, C.; Lazaro, E.; Manrubia, S.C. Quasispecies dynamics and RNA virus extinction. *Virus Res.* **2005**, *107*, 129–139. [CrossRef]
2. Domingo, E. Virus as Populations. Composition, Complexity, Dynamics and Biological Implications. In *Virus as Populations*, 1st ed.; Academic Press: Cambridge, MA, USA; Elsevier: London, UK, 2016; pp. 1–412.
3. Gregori, J.; Salicru, M.; Domingo, E.; Sanchez, A.; Esteban, J.I.; Rodriguez-Frias, F.; Quer, J. Inference with viral quasispecies diversity indices: Clonal and NGS approaches. *Bioinformatics* **2014**, *30*, 1104–1111. [CrossRef]
4. Gregori, J.; Perales, C.; Rodriguez-Frias, F.; Esteban, J.I.; Quer, J.; Domingo, E. Viral quasispecies complexity measures. *Virology* **2016**, *493*, 227–237. [CrossRef] [PubMed]
5. Vellas, C.; Del Bello, A.; Debard, A.; Steinmeyer, Z.; Tribaudeau, L.; Ranger, N.; Jeanne, N.; Martin-Blondel, G.; Delobel, P.; Kamar, N.; et al. Influence of treatment with neutralizing monoclonal antibodies on the SARS-CoV-2 nasopharyngeal load and quasispecies. *Clin. Microbiol. Infect. Off. Publ. Eur. Soc. Clin. Microbiol. Infect. Dis.* **2022**, *28*, 139.e5–139.e8. [CrossRef] [PubMed]
6. Perales, C.; Martín, V.; Ruiz-Jarabo, C.M.; Domingo, E. Monitoring sequence space as a test for the target of selection in viruses. *J. Mol. Biol.* **2005**, *345*, 451–459. [CrossRef] [PubMed]
7. Gregori, J.; Soria, M.E.; Gallego, I.; Guerrero-Murillo, M.; Esteban, J.I.; Quer, J.; Perales, C.; Domingo, E. Rare haplotype load as marker for lethal mutagenesis. *PLoS ONE* **2018**, *13*, e0204877. [CrossRef]
8. Perales, C.; Gallego, I.; de Ávila, A.I.; Soria, M.E.; Gregori, J.; Quer, J.; Domingo, E. The increasing impact of lethal mutagenesis of viruses. *Future Med. Chem.* **2019**, *11*, 1645–1657. [CrossRef]
9. Nakano, K.; Shiroma, A.; Shimoji, M.; Tamotsu, H.; Ashimine, N.; Ohki, S.; Shinzato, M.; Minami, M.; Nakanishi, T.; Teruya, K.; et al. Advantages of genome sequencing by long-read sequencer using SMRT technology in medical area. *Hum. Cell* **2017**, *30*, 149–161. [CrossRef]
10. Bull, R.A.; Eltahla, A.A.; Rodrigo, C.; Koekkoek, S.M.; Walker, M.; Pirozyan, M.R.; Betz-Stablein, B.; Toepfer, A.; Laird, M.; Oh, S.; et al. A method for near full-length amplification and sequencing for six hepatitis C virus genotypes. *BMC Genom.* **2016**, *17*, 247. [CrossRef] [PubMed]
11. Dilernia, D.A.; Chien, J.T.; Monaco, D.C.; Brown, M.P.S.; Ende, Z.; Deymier, M.J.; Yue, L.; Paxinos, E.E.; Allen, S.; Tirado-Ramos, A.; et al. Multiplexed highly-accurate DNA sequencing of closely-related HIV-1 variants using continuous long reads from single molecule, real-time sequencing. *Nucleic Acids Res.* **2015**, *43*, e129. [CrossRef] [PubMed]
12. Stoler, N.; Nekrutenko, A. Sequencing error profiles of Illumina sequencing instruments. *NAR Genom. Bioinforma.* **2021**, *3*, lqab019. [CrossRef]

13. Amarasinghe, S.L.; Su, S.; Dong, X.; Zappia, L.; Ritchie, M.E.; Gouil, Q. Opportunities and challenges in long-read sequencing data analysis. *Genome Biol.* **2020**, *21*, 30. [CrossRef]
14. Wenger, A.M.; Peluso, P.; Rowell, W.J.; Chang, P.C.; Hall, R.J.; Concepcion, G.T.; Ebler, J.; Fungtammasan, A.; Kolesnikov, A.; Olson, N.D.; et al. Accurate circular consensus long-read sequencing improves variant detection and assembly of a human genome. *Nat. Biotechnol.* **2019**, *37*, 1155–1162. [CrossRef] [PubMed]
15. Hill, M.O. Diversity and Evenness: A Unifying Notation and Its Consequences. *Ecology* **1973**, *54*, 427–432. [CrossRef]
16. Carcereny, A.; Martínez-Velázquez, A.; Bosch, A.; Allende, A.; Truchado, P.; Cascales, J.; Romalde, J.L.; Lois, M.; Polo, D.; Sánchez, G.; et al. Monitoring Emergence of the SARS-CoV-2 B.1.1.7 Variant through the Spanish National SARS-CoV-2 Wastewater Surveillance System (VATar COVID-19). *Environ. Sci. Technol.* **2021**, *55*, 11756–11766. [CrossRef]
17. Twist Synthetic SARS-CoV-2 RNA Control 2 MN908947.3. Available online: https://www.twistbioscience.com/es/resources/product-sheet/twist-synthetic-sars-cov-2-rna-controls (accessed on 12 November 2022).
18. Cubero, M.; Gregori, J.; Esteban, J.I.; Garcia-Cehic, D.; Bes, M.; Perales, C.; Domingo, E.; Rodriguez-Frias, F.; Sauleda, S.; Casillas, R.; et al. Identification of host and viral factors involved in a dissimilar resolution of a hepatitis C virus infection. *Liver Int.* **2014**, *34*, 896–906. [CrossRef] [PubMed]
19. Baccam, P.; Thompson, R.J.; Fedrigo, O.; Carpenter, S.; Cornette, J.L. PAQ: Partition Analysis of Quasispecies. *Bioinformatics* **2001**, *17*, 16–22. [CrossRef] [PubMed]
20. Töpfer, A.; Marschall, T.; Bull, R.A.; Luciani, F.; Schönhuth, A.; Beerenwinkel, N. Viral quasispecies assembly via maximal clique enumeration. *PLoS Comput. Biol.* **2014**, *10*, e1003515. [CrossRef]
21. Skums, P.; Zelikovsky, A.; Singh, R.; Gussler, W.; Dimitrova, Z.; Knyazev, S.; Mandric, I.; Ramachandran, S.; Campo, D.; Jha, D.; et al. QUENTIN: Reconstruction of disease transmissions from viral quasispecies genomic data. *Bioinformatics* **2018**, *34*, 163–170. [CrossRef]
22. Ahn, S.; Ke, Z.; Vikalo, H. Viral quasispecies reconstruction via tensor factorization with successive read removal. *Bioinformatics* **2018**, *34*, i23–i31. [CrossRef] [PubMed]
23. Henningsson, R.; Moratorio, G.; Bordería, A.V.; Vignuzzi, M.; Fontes, M. DISSEQT-DIStribution-based modeling of SEQuence space Time dynamics. *Virus Evol.* **2019**, *5*, vez028. [CrossRef] [PubMed]
24. Beerenwinkel, N.; Zagordi, O. Ultra-deep sequencing for the analysis of viral populations. *Curr.Opin.Virol.* **2011**, *1*, 413–418. [CrossRef] [PubMed]
25. Lorenzo-Redondo, R.; Delgado, S.; Morán, F.; Lopez-Galindez, C. Realistic three dimensional fitness landscapes generated by self organizing maps for the analysis of experimental HIV-1 evolution. *PLoS ONE* **2014**, *9*, e88579. [CrossRef] [PubMed]
26. Delgado, S.; Perales, C.; García-Crespo, C.; Soria, M.E.; Gallego, I.; de Ávila, A.I.; Martínez-González, B.; Vázquez-Sirvent, L.; López-Galíndez, C.; Morán, F.; et al. A Two-Level, Intramutant Spectrum Haplotype Profile of Hepatitis C Virus Revealed by Self-Organized Maps. *Microbiol. Spectr.* **2021**, *9*, e0145921. [CrossRef] [PubMed]
27. Gregori, J.; Cortese, M.F.; Piñana, M.; Campos, C.; Garcia-Cehic, D.; Andrés, C.; Abril, J.F.; Codina, M.G.; Rando, A.; Esperalba, J.; et al. Host-dependent editing of SARS-CoV-2 in COVID-19 patients. *Emerg. Microbes Infect.* **2021**, *10*, 1777–1789. [CrossRef]
28. Martínez-González, B.; Soria, M.E.; Vázquez-Sirvent, L.; Ferrer-Orta, C.; Lobo-Vega, R.; Mínguez, P.; de la Fuente, L.; Llorens, C.; Soriano, B.; Ramos-Ruíz, R.; et al. SARS-CoV-2 Mutant Spectra at Different Depth Levels Reveal an Overwhelming Abundance of Low Frequency Mutations. *Pathogens* **2022**, *11*, 662. [CrossRef]
29. De Avila, A.I.; Gallego, I.; Soria, M.E.; Gregori, J.; Quer, J.; Ignacio Esteban, J.; Rice, C.M.; Domingo, E.; Perales, C. Lethal mutagenesis of hepatitis C virus induced by favipiravir. *PLoS ONE* **2016**, *11*, e0164691. [CrossRef]
30. Perales, C.; Agudo, R.; Tejero, H.; Manrubia, S.C.; Domingo, E. Potential benefits of sequential inhibitor-mutagen treatments of RNA virus infections. *PLoS Pathog.* **2009**, *5*, e1000658. [CrossRef]
31. Githaka, J.M. Molnupiravir Does Not Induce Mutagenesis in Host Lung Cells during SARS-CoV-2 Treatment. *Bioinform. Biol. Insights* **2022**, *16*, 11779322221085076. [CrossRef]
32. Gordon, C.J.; Tchesnokov, E.P.; Schinazi, R.F.; Götte, M. Molnupiravir promotes SARS-CoV-2 mutagenesis via the RNA template. *J. Biol. Chem.* **2021**, *297*, 100770. [CrossRef] [PubMed]
33. Cameron, C.E.; Castro, C. The mechanism of action of ribavirin: Lethal mutagenesis of RNA virus genomes mediated by the viral RNA-dependent RNA polymerase. *Curr. Opin. Infect. Dis.* **2001**, *14*, 757–764. [CrossRef] [PubMed]
34. Dietz, J.; Schelhorn, S.-E.; Fitting, D.; Mihm, U.; Susser, S.; Welker, M.-W.; Füller, C.; Däumer, M.; Teuber, G.; Wedemeyer, H.; et al. Deep sequencing reveals mutagenic effects of ribavirin during monotherapy of hepatitis C virus genotype 1-infected patients. *J. Virol.* **2013**, *87*, 6172–6181. [CrossRef]
35. Cuevas, J.M.; González-Candelas, F.; Moya, A.; Sanjuán, R. Effect of ribavirin on the mutation rate and spectrum of hepatitis C virus in vivo. *J. Virol.* **2009**, *83*, 5760–5764. [CrossRef]
36. Todt, D.; Meister, T.L.; Steinmann, E. Hepatitis E virus treatment and ribavirin therapy: Viral mechanisms of nonresponse. *Curr. Opin. Virol.* **2018**, *32*, 80–87. [CrossRef]
37. Todt, D.; Gisa, A.; Radonic, A.; Nitsche, A.; Behrendt, P.; Suneetha, P.V.; Pischke, S.; Bremer, B.; Brown, R.J.P.; Manns, M.P.; et al. In vivo evidence for ribavirin-induced mutagenesis of the hepatitis E virus genome. *Gut* **2016**, *65*, 1733–1743. [CrossRef]
38. Debing, Y.; Gisa, A.; Dallmeier, K.; Pischke, S.; Bremer, B.; Manns, M.; Wedemeyer, H.; Suneetha, P.V.; Neyts, J. A mutation in the hepatitis E virus RNA polymerase promotes its replication and associates with ribavirin treatment failure in organ transplant recipients. *Gastroenterology* **2014**, *147*, 1006–1008. [CrossRef] [PubMed]

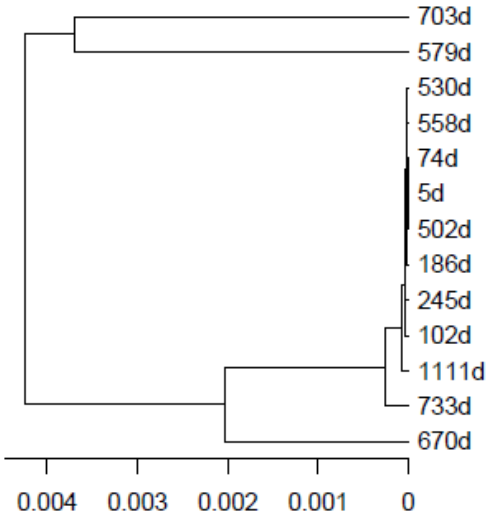
39. Lhomme, S.; Kamar, N.; Nicot, F.; Ducos, J.; Bismuth, M.; Garrigue, V.; Petitjean-Lecherbonnier, J.; Ollivier, I.; Alessandri-Gradt, E.; Goria, O.; et al. Mutation in the Hepatitis E Virus Polymerase and Outcome of Ribavirin Therapy. *Antimicrob. Agents Chemother.* **2015**, *60*, 1608–1614. [CrossRef]
40. Debing, Y.; Ramière, C.; Dallmeier, K.; Piorkowski, G.; Trabaud, M.-A.; Lebossé, F.; Scholtès, C.; Roche, M.; Legras-Lachuer, C.; de Lamballerie, X.; et al. Hepatitis E virus mutations associated with ribavirin treatment failure result in altered viral fitness and ribavirin sensitivity. *J. Hepatol.* **2016**, *65*, 499–508. [CrossRef] [PubMed]
41. ARTIC Network. Available online: https://artic.network/ncov-2019 (accessed on 12 November 2022).
42. Andrés, C.; Garcia-Cehic, D.; Gregori, J.; Piñana, M.; Rodriguez-Frias, F.; Guerrero-Murillo, M.; Esperalba, J.; Rando, A.; Goterris, L.; Codina, M.G.; et al. Naturally occurring SARS-CoV-2 gene deletions close to the spike S1/S2 cleavage site in the viral quasispecies of COVID19 patients. *Emerg. Microbes Infect.* **2020**, *9*, 1900–1911. [CrossRef]
43. Gallego, I.; Gregori, J.; Soria, M.E.; Garcia-Crespo, C.; Garcia-Alvarez, M.; Gomez-Gonzalez, A.; Valiergue, R.; Gomez, J.; Esteban, J.I.; Quer, J.; et al. Resistance of high fitness hepatitis C virus to lethal mutagenesis. *Virology* **2018**, *523*, 100–109. [CrossRef]
44. Gallego, I.; Sheldon, J.; Moreno, E.; Gregori, J.; Quer, J.; Esteban, J.I.; Rice, C.M.; Domingo, E.; Perales, C. Barrier-independent, fitness-associated differences in sofosbuvir efficacy against hepatitis c virus. *Antimicrob. Agents Chemother.* **2016**, *60*, 3786–3793. [CrossRef]
45. Marukian, S.; Jones, C.T.; Andrus, L.; Evans, M.J.; Ritola, K.D.; Charles, E.D.; Rice, C.M.; Dustin, L.B. Cell culture-produced hepatitis C virus does not infect peripheral blood mononuclear cells. *Hepatology* **2008**, *48*, 1843–1850. [CrossRef] [PubMed]
46. Perales, C.; Beach, N.M.; Gallego, I.; Soria, M.E.; Quer, J.; Esteban, J.I.; Rice, C.; Domingo, E.; Sheldon, J. Response of hepatitis C virus to long-term passage in the presence of alpha interferon: Multiple mutations and a common phenotype. *J. Virol.* **2013**, *87*, 7593–7607. [CrossRef] [PubMed]
47. Moreno, E.; Gallego, I.; Gregori, J.; Lucía-Sanz, A.; Soria, M.E.; Castro, V.; Beach, N.M.; Manrubia, S.; Quer, J.; Esteban, J.I.; et al. Internal Disequilibria and Phenotypic Diversification during Replication of Hepatitis C Virus in a Noncoevolving Cellular Environment. *J. Virol.* **2017**, *91*, e02505-16. [CrossRef] [PubMed]
48. Magoc, T.; Salzberg, S.L. FLASH: Fast length adjustment of short reads to improve genome assemblies. *Bioinformatics* **2011**, *27*, 2957–2963. [CrossRef]
49. Chao, A.; Gotelli, N.J.; Hsieh, T.C.; Sander, E.L.; Ma, K.H.; Colwell, R.K.; Ellison, A.M. Rarefaction and extrapolation with Hill numbers: A framework for sampling and estimation in species diversity studies. *Ecol. Monogr.* **2014**, *84*, 45–67. [CrossRef]
50. Gregori, J.; Esteban, J.I.; Cubero, M.; Garcia-Cehic, D.; Perales, C.; Casillas, R.; Alvarez-Tejado, M.; Rodriguez-Frias, F.; Guardia, J.; Domingo, E.; et al. Ultra-deep pyrosequencing (UDPS) data treatment to study amplicon HCV minor variants. *PLoS ONE* **2013**, *8*, e0083361. [CrossRef] [PubMed]
51. Nei, M. *Molecular Evolutionary Genetics*; Columbia University Press: New York, NY, USA, 1987; ISBN 9780231886710.
52. Grishin, V.N.; Grishin, N. V Euclidian space and grouping of biological objects. *Bioinformatics* **2002**, *18*, 1523–1534. [CrossRef] [PubMed]
53. Fitch, W.M. An improved method of testing for evolutionary homology. *J. Mol. Biol.* **1966**, *16*, 9–16. [CrossRef] [PubMed]
54. Grantham, R. Amino acid difference formula to help explain protein evolution. *Science* **1974**, *185*, 862–864. [CrossRef]
55. Yue, J.C.; Clayton, M.K. A Similarity Measure Based on Species Proportions. *Commun. Stat.-Theory Methods* **2005**, *34*, 2123–2131. [CrossRef]
56. R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2019.
57. Pages, H.; Aboyoun, P.; Gentleman, R.; DebRoy, S. Biostrings: String Objects Representing Biological Sequences, and Matching Algorithms, R package 2.38.4; 2012. Available online: https://bioc.ism.ac.jp/packages/3.2/bioc/html/Biostrings.html (accessed on 12 November 2022).
58. Morgan, M.; Anders, S.; Lawrence, M.; Aboyoun, P.; Pages, H.; Gentleman, R. ShortRead: A bioconductor package for input, quality assessment and exploration of high-throughput sequence data. *Bioinformatics* **2009**, *25*, 2607–2608. [CrossRef]
59. Guerrero-Murillo, M.; Gregori, J. QSutils: Quasispecies Diversity. R Package Version 1.0.0. Available online: https://bioconductor.org/packages/release/bioc/html/QSutils.html_2018 (accessed on 22 November 2022).
60. Gentleman, R.C.; Carey, V.J.; Bates, D.M.; Bolstad, B.; Dettling, M.; Dudoit, S.; Ellis, B.; Gautier, L.; Ge, Y.; Gentry, J.; et al. Bioconductor: Open software development for computational biology and bioinformatics. *Genome Biol.* **2004**, *5*, R80. [CrossRef]
61. Paradis, E.; Schliep, K. ape 5.0: An environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* **2019**, *35*, 526–528. [CrossRef] [PubMed]
62. Wickham, H. Welcome to Master the Tidyverse. *J. Open Source Softw.* **2019**, *4*. [CrossRef]
63. Valero-Mora, P.M. ggplot2: Elegant Graphics for Data Analysis. *J. Stat. Soft. Book Rev.* **2010**, *35*, 1–3. [CrossRef]

4.1.3 Article 1: Supplementary Material

Quasispecies Fitness Partition to Characterize the Molecular Status of a Viral Population. Negative Effect of Early Ribavirin Discontinuation in a Chronically Infected HEV Patient



(a)



(b)

Figure S1: UPGMA tree of the master phenotypes based on Grantham amino acid distances (a), and quasispecies tree based on the Da population distances (b), taking the top 20 phenotypes in each sample.

1

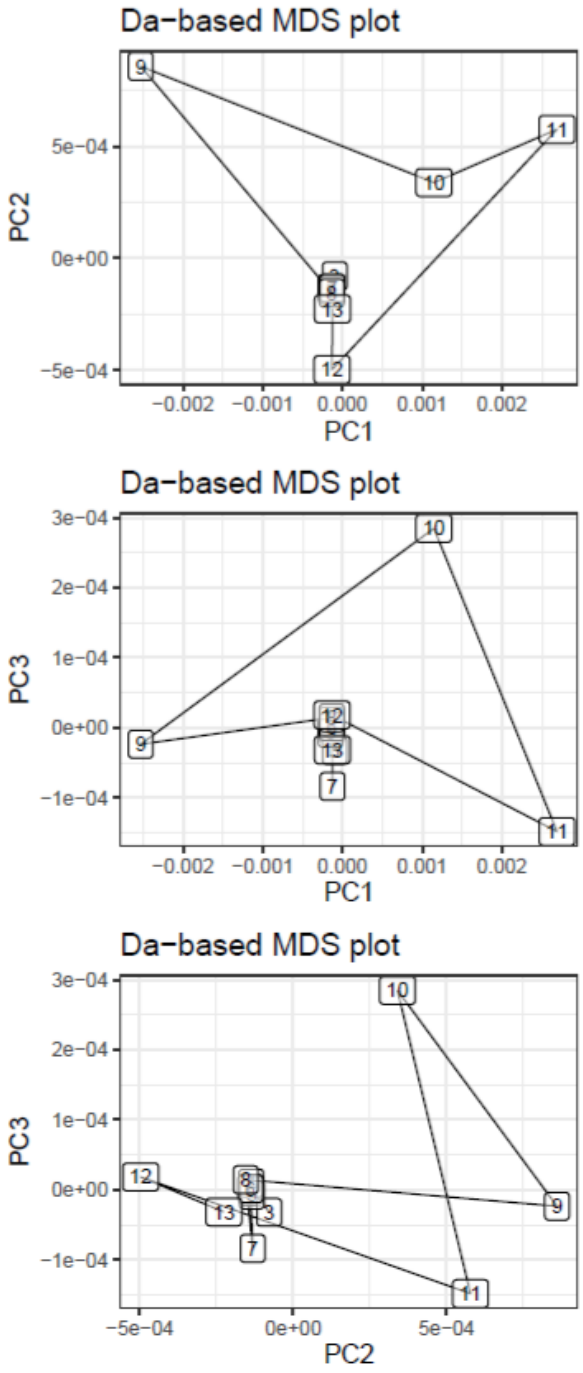


Figure S2: Multidimensional scaling plot of phenotype distances between quasispecies, taking the top 20 phenotypes in each sample.

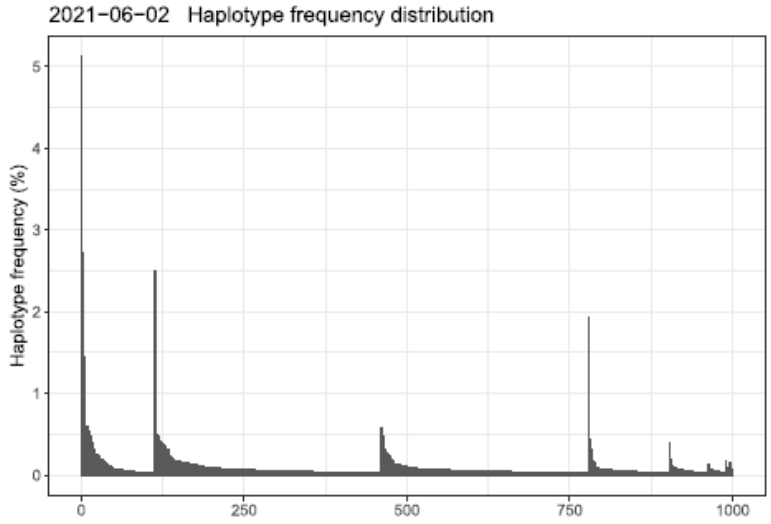


Figure S3: Montserrat plot with the distribution of the 1000 most abundant haplotypes in the last sample (1111 days since diagnosis, 2 June 2020).

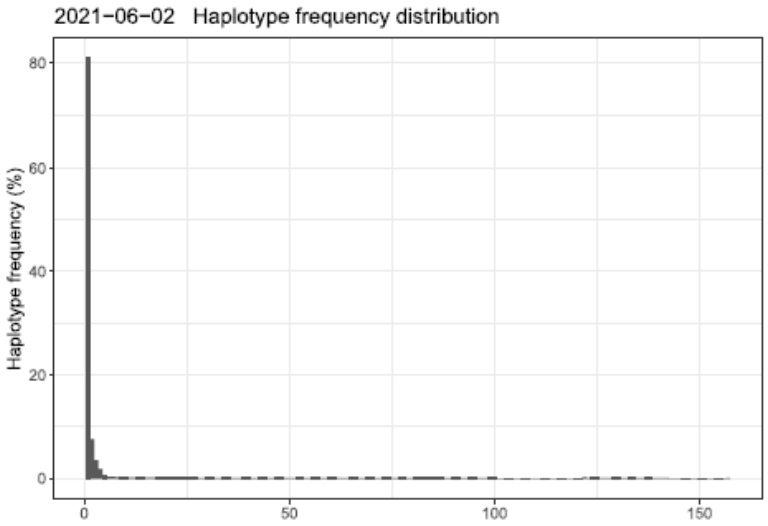


Figure S4: Montserrat plot with the distribution of all phenotypes in the last sample (1111 days since diagnosis, 2 June 2020).

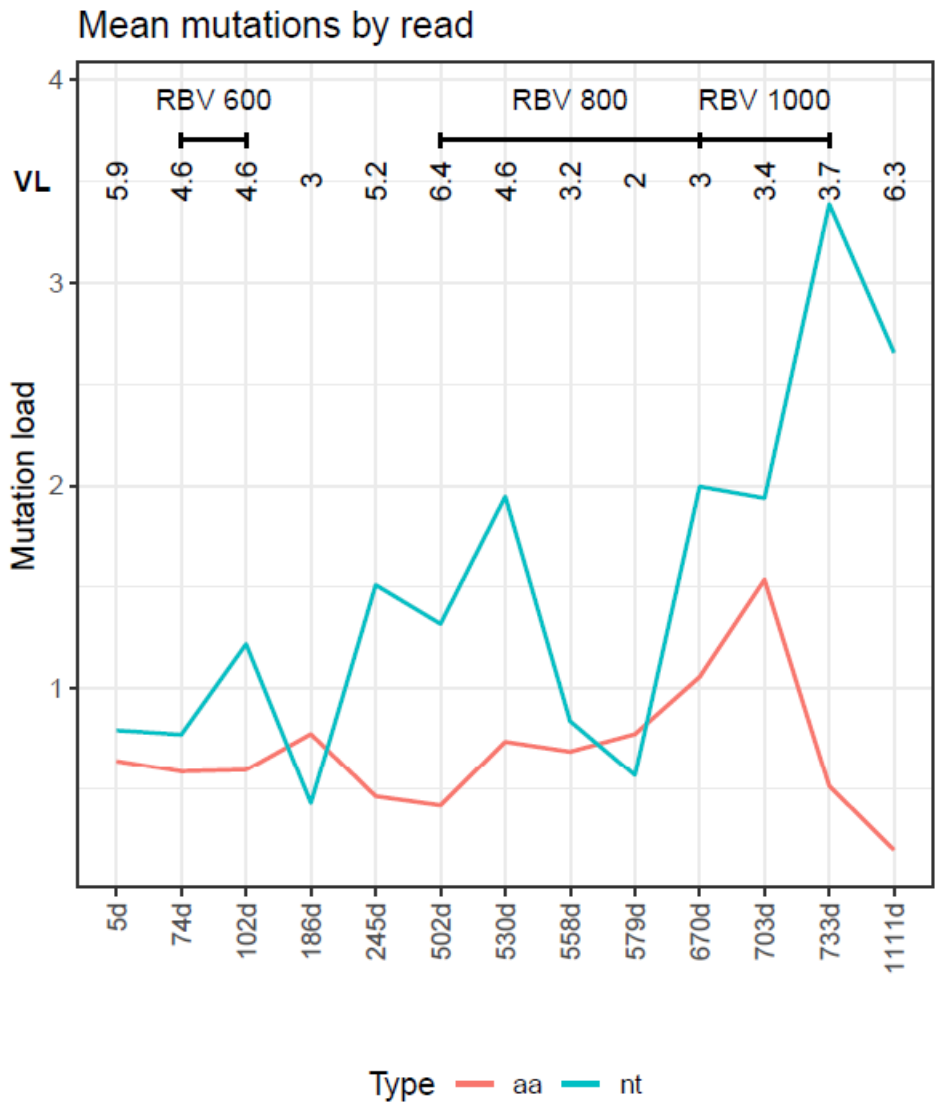


Figure S5: Mean number of substitutions per read with respect to the master haplotype, at the nucleotide level (turquoise), and mean number of mutations per read with respect to the master phenotype at the amino acid level (orange).

Table S1. Primers Forward (FW) and Reverse (RV) for hepatitis E virus (HEV), hepatitis C virus (HCV) and SARS-CoV-2. FW.pos and REV.pos are the primer positions. FW.tpos and RV.tpos are the trimming positions. AmplLen indicates the amplicon length after trimming primers.

Code	Pathogen	Region	Primer.FW	Primer.RV	FW.pos	RV.pos	AmplLen
gHEV	HEV	ORF2	GTCGTCTCAGCCAATGGCGAGCC	CASARAANGTCTTNGARTACTGCT	6323	6734	363
A1	HCV	NS5B	TGGTCTACTTGCTCCGAGGAGG	TTGGCCCCGAATCCATACTTG	8237	8595	312
A2	HCV	NS5B	CTTGGAGGAGGCGTGCCAGTTG	GATACTCCACCCGTTGGGC	8530	8888	318
A3	HCV	NS5B	AGCTTCCTCAGGCGGTAATGG	TCAGGTTCCGCTCGTCCTCCTC	8820	9287	423
A71	SARS-CoV-2	ORF1ab	ACAAATCCAATTCAGTTGTCTTCCTATTC	TGGAAAAGAAAGGTAAGAACAAGTCCT	21358	21743	330
A72	SARS-CoV-2	S	ACACGTGGTGTTTATTACCCTGAC	ACTCTGAACTCACTTTCCATCCAAC	21659	22038	331
A73	SARS-CoV-2	S	CAATTTTGTAATGATCCATTTTTGGGTGT	CACCAGCTGTCCAACCTGAAGA	21962	22346	334
A74	SARS-CoV-2	S	ACATCACTAGGTTTCAAACTTTACTTGC	GCAACACAGTTGCTGATTCTCTTC	22263	22650	336
A75	SARS-CoV-2	S	AGAGTCCAACCAACAGAATCTATTGT	ACCACCAACCTTAGAATCAAGATTGT	22517	22903	335
A76	SARS-CoV-2	S	GGGCAAACTGGAAAGATTGCTGA	ACCTGTGCCTGTTAAACCATTGA	22799	23212	368
A77	SARS-CoV-2	S	CCAGCAACTGTTTGTGGACCTA	CAGCCCCTATTAAACAGCCTGC	23123	23522	356
A78	SARS-CoV-2	S	CAACTTACTCCTACTTGGCGTGT	TGTGTACAAAAACTGCCATATTGCA	23444	23847	356
A79	SARS-CoV-2	S	GTGGTGATTCAACTGAATGCAGC	CATTTCATCTGTGAGCAAAGGTGG	23790	24169	333
A80	SARS-CoV-2	S	TTGCCTTGGTGATATTGCTGCT	TGGAGCTAAGTTGTTTAACAAGCG	24079	24467	343
A81	SARS-CoV-2	S	GCACTTGGAAAACTTCAAGATGTGG	GTGAAGTTCTTTTCTTGTGCAGGG	24392	24789	349
A82	SARS-CoV-2	S	GGGCTATCATCTTATGTCCTTCCCT	TGCCAGAGATGTCACCTAAATCAA	24697	25076	331
A83	SARS-CoV-2	S	TCCTTTGCAACCTGAATTAGACTCA	TTTGACTCCTTTGAGCACTGGC	24979	25369	344
A84	SARS-CoV-2	S	TGCTGTAGTTGTCTCAAGGGCT	AGGTGTGAGTAAACTGTTACAAACAAC	25280	25673	345

4.2 Study 2

In-Host HEV Quasispecies Evolution Shows the Limits of Mutagenic Antiviral Treatments.

4.2.1 Article 2 Summary

4.2.1.1 Introduction

HEV is a major global cause of acute viral hepatitis, with approximately 20 million infections and 44000 deaths annually. While usually self-limiting, the infection can progress to a chronic stage in immunocompromised patients, such as organ transplant recipients. Current clinical guidelines recommend a three-month RBV monotherapy as the standard treatment for persistent cases. When relapse, only a second RBV treatment is defined, with the alternative for PEG-Interferon only in the case of liver transplant patients. HEV is a highly variable RNA virus circulating as quasispecies. Since RBV could have mutagenic action, its mechanism of action is hypothesized to involve lethal mutagenesis, aiming to push the virus towards extinction by increasing its mutation rate beyond a sustainable threshold. However, the rising incidence of treatment failure and the emergence of resistance mutation highlight a critical need to understand how the quasispecies evolve under long-term drug pressure.

4.2.1.2 Objectives

1. To describe in-host quasispecies evolution of HEV in patients with long-term chronic infection.
2. To analyse and compare population dynamics at the genotypic (nucleotide) and phenotypic (amino acid) levels.
3. To assess the impact of RBV treatment on quasispecies structure.

4.2.1.3 Study design

- The study is based in a longitudinal clinical follow-up of a 62-year-old male patient with lung a kidney transplantation who developed chronic HEV during 6 years:
    - 12 sequential plasma samples were collected at various time points across the infection course including periods of treatment and end of treatment (EOT)
- Viral RNA was extracted and subjected to RT-PCR amplification for a conserved ORF2 (core) region.
- Deep sequencing was performed using the Illumina MiSeq platform (amplicons of 300-500bp).
- The evolution of the population was analysed at both genotypic and phenotypic levels by using QFF, HNP and UPGMA trees.

4.2.1.4 Results according to objectives

4.2.1.4.1 Study 2. Objective 1: To describe in-host quasispecies evolution of HEV in patients with long-term chronic infection.

The QFF study revealed that the master haplotype was maintained at medium frequencies (24.7%-44.3%) until day 1162. After the second EOT (no RNA negativization), a new master haplotype appeared in day 1358 with a frequency of 59.7%. In the four final samples of the study (days 1414 to day 2135), viral population became in an unstructured cloud of genomes, as previously reported in the previous paper of this compendium, where the new master haplotypes were represented in low proportion (4.1%-6.1%). At this moment, an important emergent haplotype fraction appears while the combination of both RHL is near the 75% of all que quasispecies (**Figure 13,** corresponding to figure 2 in Article 2).

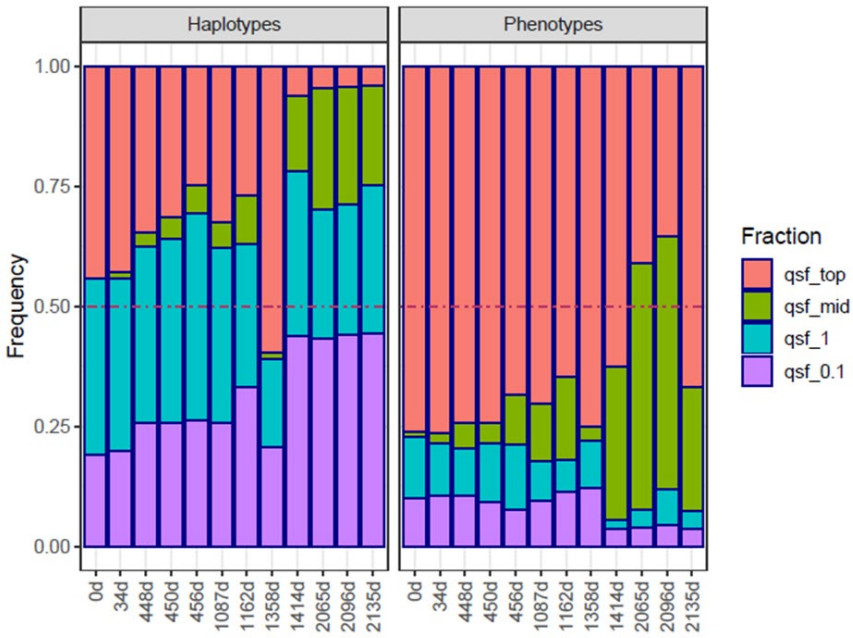


Figure 13. Quasispecies Fitness Fractions for each sample, at the genetic level on the left, and at the phenotypic level on the right. qsf_top: fraction of reads belonging to the master; qsf_mid: fraction of reads belonging to emerging haplotypes; qsf_1: fraction of reads of rare haplotypes; qsf_0.1: fraction of reads of very rare and defective haplotypes. The dash-dot line at 0.5 helps in visualizing the masters of either type with a frequency above a 50%.

4.2.1.4.2 Study 2. Objective 2: To analyse and compare population dynamics at the genotypic (nucleotide) and phenotypic (amino acid) levels.

In agreement with Study 1, here we found in a new patient, that there was a profound discrepancy between the quasispecies analysed at the genotypic and at the phenotypic levels. While genotypes revealed a complex quasispecies with multiple haplotypes competing with the master, phenotypic QFF showed the dominance of the master haplotype, representing approximately 75% of the quasispecies. However, in the last four samples, as observed for genotypes, there was a substantial increase in the frequency of emergent haplotypes.

In addition, in the last three samples, although the master haplotypes differed phylogenetically, all shared the same master phenotype, with a maximum

prevalence of 66.86% in the last sample. Moreover, the genetic similarity among phenotypes remained relatively high (0.5), suggesting that the virus retained its functionality despite the nucleotide mutations.

4.2.1.4.3 Study 2. Objective 3: To assess the impact of RBV treatment on quasispecies structure.

RBV treatment increased significantly the proportions of RHL and the complexity of the mutational spectra. There were high multiplicity mutants, with even 15 mutations in the same amplicon under prolongated treatment (**Figure 14,** corresponding to figure 4 in Article 2). However, the amino acid pattern of substitutions revealed that the viral population moved to the production of neutral mutations, since the more prevalent changes were conservative. Finally, the transitions associated to RBV effect (C → T/C and G → A/G) were notably prevalent in the last four samples coinciding with the highest complexity of the quasispecies (**Figure 15,** corresponding to figure 6 in Article 2).

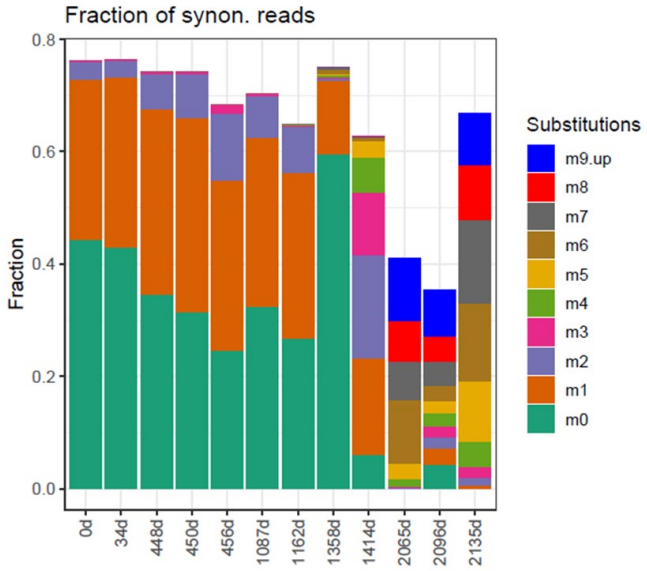


Figure 14. Aggregation of reads of synonymous haplotypes to the master phenotype, according to the number of substitutions with respect to the master haplotype in each sample. m0: fraction of reads for the master, m1: fraction of reads with one substitution with respect to the master haplotype, and so on.

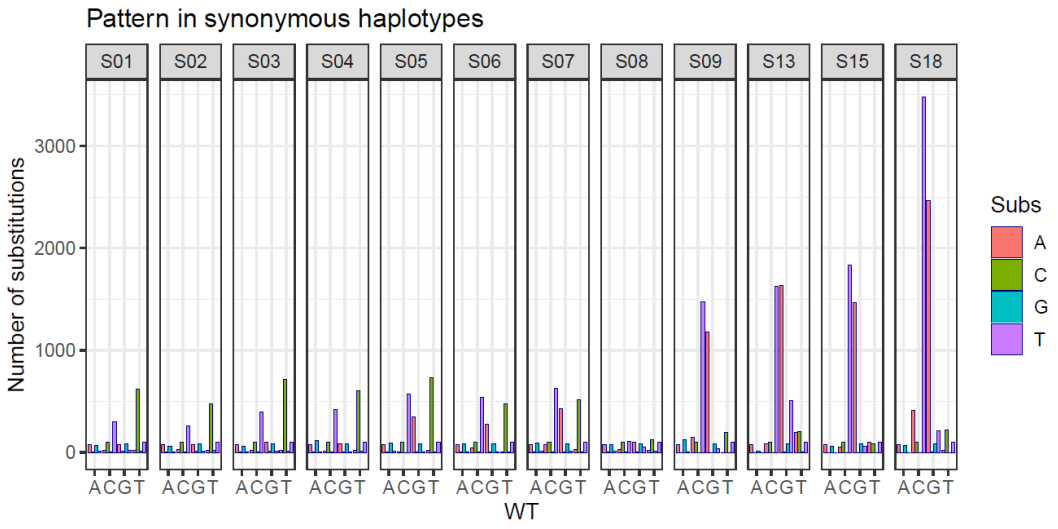


Figure 15. Substitution pattern in synonymous haplotypes to the wild type (consensus synonymous haplotype in each sample).

In summary, the results obtained in Studies 1 and 2 included in this thesis, involving two chronically infected patients who failed sequential RBV monotherapy treatments, showed that RBV administration in the context of high viral loads leads to an unstructured quasispecies at the nucleotide level, accumulating nucleotide substitutions while preserving protein functionality, primarily through synonymous mutations. Consequently, the number of haplotypes increases, ultimately generating a highly complex viral quasispecies at the nucleotide level, that allows the proliferation of high-fitness variants and may compromise subsequent treatments.

4.2.1.5 Conclusions

- There is a profound difference in how quasispecies responds to RBV at the genotypic versus phenotypic level. At the nucleotide level, the quasispecies become highly unstructured, showing a drastic reduction in the master sequence and a rise in rare haplotypes. However, at the amino acid level, the master phenotype remains highly prevalent and conserved, suggesting the virus maintains its functional fitness.
- Ribavirin treatment, particularly at high viral loads, induces the selection of a diverse set of synonymous haplotypes. These synonymous variants act as reservoir of functional genomes that could replace the master sequence.
- One clinical application of our results is that, during chronic HEV infection, initial RBV treatment in monotherapy should be continued even after serum RNA becomes undetectable for several weeks to prevent relapse. Otherwise, any relapse may promote the selection of fitness-enhancing mutations, ultimately leading to treatment failure.
- Another clinical implication is that relapse following ineffective ribavirin treatment, which fails to achieve viral extinction, can increase viral diversity and carry the risk of selecting fitness-enhancing mutations. Therefore, mutagenic therapy may be most effective when combined with other antiviral inhibitors to minimize viral replication.

4.2.2 Article 2

In-Host HEV Quasispecies Evolution Shows the Limits of Mutagenic Antiviral Treatments.

Sergi Colomer-Castell†, Josep Gregori†, Damir Garcia-Cehic, Mar Riveiro-Barciela, Maria Buti, Ariadna Rando-Segura, Judit Vico-Romero, Carolina Campos, Marta Ibañez-Lligoña, Caroline Melanie Adombi, Maria Francesca Cortese, David Tabernero, Juan Ignacio Esteban, Francisco Rodriguez-Frias and Josep Quer*

International Journal of
Molecular Sciences

MDPI

Article

In-Host HEV Quasispecies Evolution Shows the Limits of Mutagenic Antiviral Treatments

Sergi Colomer-Castell [1,2,3,†], Josep Gregori [1,*,†], Damir Garcia-Cehic [1,2], Mar Riveiro-Barciela [1,2,4], Maria Buti [1,2,4], Ariadna Rando-Segura [2,5], Judit Vico-Romero [1], Carolina Campos [1,2,3], Marta Ibañez-Lligoña [1,4], Caroline Melanie Adombi [1,6], Maria Francesca Cortese [2,7], David Tabernero [2,7], Juan Ignacio Esteban [1,2,4], Francisco Rodriguez-Frias [2,7] and Josep Quer [1,2,3,*]

1 Liver Diseases-Viral Hepatitis, Liver Unit, Vall d'Hebron Institut de Recerca (VHIR), Vall d'Hebron Hospital Universitari, Vall d'Hebron Barcelona Hospital Campus, Passeig Vall d'Hebron 119-129, 08035 Barcelona, Spain; sergi.colomer@vhir.org (S.C.-C.); damir.garcia@vhir.org (D.G.-C.); mar.riveiro@gmail.com (M.R.-B.); maria.buti@vallhebron.cat (M.B.); judit.vico@vhir.org (J.V.-R.); carolina.campos@vhir.org (C.C.); marta.ibanez@vhir.org (M.I.-L.); adombicaro@yahoo.fr (C.M.A.); juanignacio.esteban@vallhebron.cat (J.I.E.)
2 Centro de Investigación Biomédica en Red de Enfermedades Hepáticas y Digestivas (CIBERehd), Instituto de Salud Carlos III, Av. Monforte de Lemos, 3-5, 28029 Madrid, Spain; ariadna.rando@vallhebron.cat (A.R.-S.); maria.cortese@vhir.org (M.F.C.); dtaberc@gmail.com (D.T.); frarodri@gmail.com (F.R.-F.)
3 Biochemistry and Molecular Biology Department, Universitat Autònoma de Barcelona (UAB), Campus de la UAB, Plaça Cívica, 08193 Bellaterra, Spain
4 Medicine Department, Universitat Autònoma de Barcelona (UAB), Campus de la UAB, Plaça Cívica, 08193 Bellaterra, Spain
5 Microbiology Department, Vall d'Hebron Institut de Recerca (VHIR), Vall d'Hebron Hospital Universitari, Vall d'Hebron Barcelona Hospital Campus, Passeig Vall d'Hebron 119-129, 08035 Barcelona, Spain
6 Institute of Agropastoral Management, University Peleforo Gon Coulibaly, Korhogo BP 1328, Côte d'Ivoire
7 Biochemistry Department, Vall d'Hebron Institut de Recerca (VHIR), Vall d'Hebron Hospital Universitari, Vall d'Hebron Barcelona Hospital Campus, Passeig Vall d'Hebron 119-129, 08035 Barcelona, Spain
* Correspondence: josep.gregori@gmail.com or josep.gregori@vhir.org (J.G.); josep.quer@vhir.org (J.Q.)
† These authors contributed equally to this work.

Abstract: Here, we report the in-host hepatitis E virus (HEV) quasispecies evolution in a chronically infected patient who was treated with three different regimens of ribavirin (RBV) for nearly 6 years. Sequential plasma samples were collected at different time points and subjected to RNA extraction and deep sequencing using the MiSeq Illumina platforms. Specifically, we RT-PCR amplified a single amplicon from the core region located in the open-reading frame 2 (ORF2). At the nucleotide level (genotype), our analysis showed an increase in the number of rare haplotypes and a drastic reduction in the frequency of the master (most represented) sequence during the period when the virus was found to be insensitive to RBV treatment. Contrarily, at the amino acid level (phenotype), our study revealed conservation of the amino acids, which is represented by a high prevalence of the master sequence. Our findings suggest that using mutagenic antivirals concomitant with high viral loads can lead to the selection and proliferation of a rich set of synonymous haplotypes that express the same phenotype. This can also lead to the selection and proliferation of conservative substitutions that express fitness-enhanced phenotypes. These results have important clinical implications, as they suggest that using mutagenic agents as a monotherapy treatment regimen in the absence of sufficiently effective viral inhibitors can result in diversification and proliferation of a highly diverse quasispecies resistant to further treatment. Therefore, such approaches should be avoided whenever possible.

Keywords: quasispecies; deep sequencing; variability; rare haplotypes; fitness; mutagens

check for updates

Citation: Colomer-Castell, S.; Gregori, J.; Garcia-Cehic, D.; Riveiro-Barciela, M.; Buti, M.; Rando-Segura, A.; Vico-Romero, J.; Campos, C.; Ibañez-Lligoña, M.; Adombi, C.M.; et al. In-Host HEV Quasispecies Evolution Shows the Limits of Mutagenic Antiviral Treatments. *Int. J. Mol. Sci.* **2023**, *24*, 17185. https://doi.org/10.3390/ijms242417185

Academic Editor: Qibin Geng

Received: 11 November 2023
Revised: 1 December 2023
Accepted: 3 December 2023
Published: 6 December 2023

1. Introduction

HEV is a major cause of acute viral hepatitis globally, particularly in low- and middle-income countries, and its incidence is on the rise in industrialized nations. According

to the WHO, approximately 20 million people are infected with HEV every year, out of which 3.3 million exhibit symptoms, and 44,000 die due to hepatic failure [1]. In Spain, the last epidemiologic study reported a 15% IgG seroprevalence in the population [2]. In addition, 1 in 3333 blood donations is positive for HEV-RNA in the Catalonia area [3]. HEV is a single-stranded positive-sense genome of 7.2 kb in length that belongs to the genus *Orthohepevirus* of the family *Hepeviridae*. HEV can be clustered genetically into eight genotypes [4]. Genotypes G1 to G4 infect humans, with G1 and G2 exclusively infecting humans through a fecal–oral route due to contamination of water supplies or food [5], whereas G3 and G4 are endemic of domestic pigs, wild boar, and deer, causing zoonotic infections in humans through consumption of uncooked or inadequately cooked processed pork meat [6–8] but also after transfusions of blood derivatives. In most cases, hepatitis E infecting humans causes acute self-limiting and asymptomatic infections that resolve within 2–8 weeks [9]. Occasionally, a serious disease known as fulminant hepatitis (acute liver failure) develops, which can be fatal [10]. In immunocompromised patients, such as solid organ transplant recipients, those receiving immunosuppressors to prevent organ rejection, patients infected by the human immunodeficiency virus (HIV) with a low CD4+ cell count, and patients receiving chemotherapy for hematological disorders G3 and G4 infection, can have a persistent infection named chronic hepatitis E [11], which is diagnosed after 3 months of continuous viremia [12].

Current clinical guidelines recommend an initial reduction in immunosuppression as the first step to manage chronic HEV, especially in solid organ-transplanted patients [9]. Nevertheless, if HEV is not eliminated, a 3-month ribavirin (RBV) monotherapy with a weight-adjusted dose or a dose based on glomerular filtration rate is the standard treatment. The sustainable virological response (SVR) has been estimated up to 78% [13,14]. When HEV relapse or the first treatment fails, a re-treatment with ribavirin for 6 months has to be carried out. However, there is no knowledge about the mechanism of action of RBV in HEV clearance, the principal hypothesis being the depletion of cellular GTP pools [15] or the lethal mutagenesis [16–18]. On the other hand, PEG-Interferon-α has shown effectiveness in HEV clearance in liver-transplanted patients, but it is contraindicated in other transplants due to its immune activation mode of action [9]. Therapeutic alternatives have been proposed, such as the combination of ribavirin with sofosbuvir, but negative results have been reported. The lack of consensus about RBV doses and treatment times, together with the apparition of mutations associated with RBV resistance, evidences the need for new antiviral treatments [19,20].

NGS has been reported as the most accurate methodology to study highly variable viruses such as HEV [16]. HEV is a is a highly variable virus, showing a rate of fixation of mutations of 1.41–1.72 $\times$ 10^{-3} substitutions/nucleotide/year [21], causing continuous production of variants during infection and generating a complex mixture of different but closely related genomes known as quasispecies [22,23]. During chronic infection, typically, a dominant strain (wild type) is detectable within the viral quasispecies along with strains that are present at lower frequencies [24]. We have recently showed that a quasispecies partition into fitness fractions (QFF, quasispecies fitness fractions) provides a valuable visualization, with biological/clinical implications, that contributes to explain the molecular changes in the composition of a quasispecies over time [16].

Here, we report the deep-sequencing study of samples sequentially collected from a HEV chronically infected patient that received three treatments with RBV. We found profound discrepancies between the nucleotide population patterns (genotype level) compared to the protein patterns (phenotype level) that have important implications in the evolution of the viral infection, disease progression, and treatment strategies, having general implications for any antiviral treatment prescription.

2. Results

In this study, we observe the cumulative effect of three different ribavirin treatment regimens, with EOTs and large periods with no treatments in between. The timeline with

interventions, samplings, and viral loads is shown in Table 1 and Supplementary Figure S1. The evolution in the quasispecies is studied at the genotype and phenotype levels with the following tools: (a) UPGMA tree of master sequences (Figure 1); (b) UPGMA tree of quasispecies (Supplementary Figure S4); (c) quasispecies structure by the method of the fitness fractions (Figure 2) and quasispecies diversity by Hill number profiles (Supplementary Figure S7); (d) evolution in the fraction of synonymous reads and haplotypes to the master phenotype (Figure 3); (e) evolution of the substitution and mutation loads in the quasispecies, corresponding to synonymous haplotypes to the master phenotype (Figure 4); (f) distribution of the top 10 haplotypes and phenotypes in each sample (Figure 5); and (g) characterization of the impact of emerging phenotypes in protein functionality (Supplementary Figure S11). Additional analyses have been carried out and are shown in the Supplementary Results (including Supplementary Figure S2, Supplementary Figure S3, Supplementary Figure S5 and Supplementary Figure S6; Supplementary Tables S1–S5).

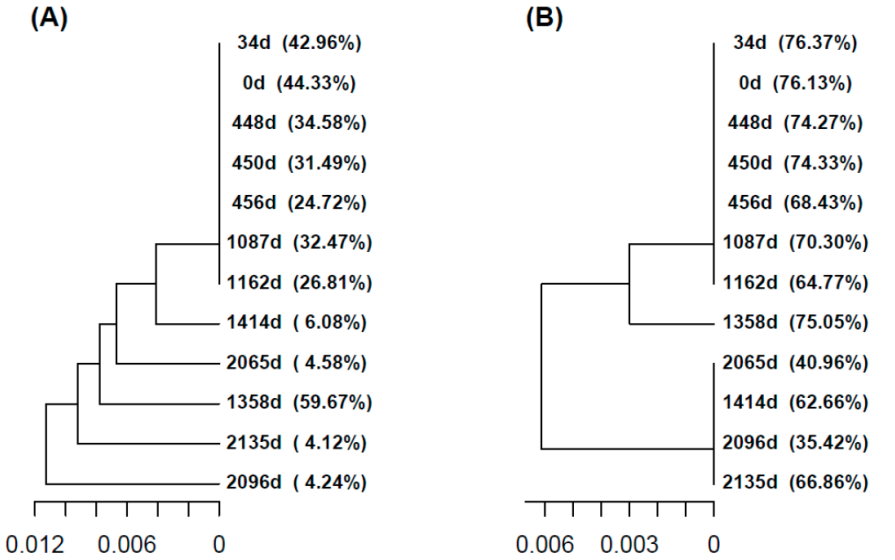


Figure 1. (**A**) UPGMA tree with the master haplotypes of all samples, at the nucleotide level. (**B**) Corresponding tree with the master phenotypes (amino acid level) of all samples in the study. Samples are labelled as days since first evidence. The observed frequencies of each sequence are expressed as percentages.

Table 1. Treatment regimens and samplings during patient disease. Periods with treatment are shown with a light-grey background. RBV: Ribavirin; EOT: End of Treatment.

Intervention *	ID [a]	Date	Days [b]	Weeks [c]	CumDays [d]	Viral Load
Infection		October 2012				
1st evidence	S01	13 November 2012	0	0.0	0	5.04×10^6
	S02	15 December 2012	34	4.9	34	6.84×10^6
RBV 200 mg		3 February 2014	413	59.0	447	
	S03	4 February 2014	1	0.1	448	4.00×10^6
	S04	6 February 2014	2	0.3	450	2.78×10^6
	S05	12 February 2014	6	0.9	456	1.17×10^6
EOT		3 May 2014	80	11.4	536	RNA−

Table 1. *Cont.*

Intervention *	ID [a]	Date	Days [b]	Weeks [c]	CumDays [d]	Viral Load
Relapse		1 January 2015	243	34.7	779	RNA+
	S06	5 November 2015	308	44.0	1087	3.56×10^6
RBV 200/400 mg	S07	19 January 2016	75	10.7	1162	1.30×10^7
EOT		5 July 2016	168	19.7	1300	RNA+
	S08	2 August 2016	28	8.3	1358	2.00×10^4
	S09	27 September 2016	56	8.0	1414	6.93×10^5
RBV 400 mg		9 April 2018	559	79.9	1973	1.27×10^7
	S13	10 July 2018	92	13.1	2065	1.51×10^5
	S15	10 August 2018	31	4.4	2096	8.22×10^4
	S18	18 September 2018	39	5.6	2135	7.50×10^4
EOT		27 November 2018	70	10.0	2205	8.20×10^4
		2 March 2019	95	13.6	2300	7.00×10^6

* Intervention includes the different treatments, samplings, and viral load monitorizing in which sample was taken. [a] Sample ID. [b] Days since previous row. [c] Weeks since previous row. [d] Days since first evidence.

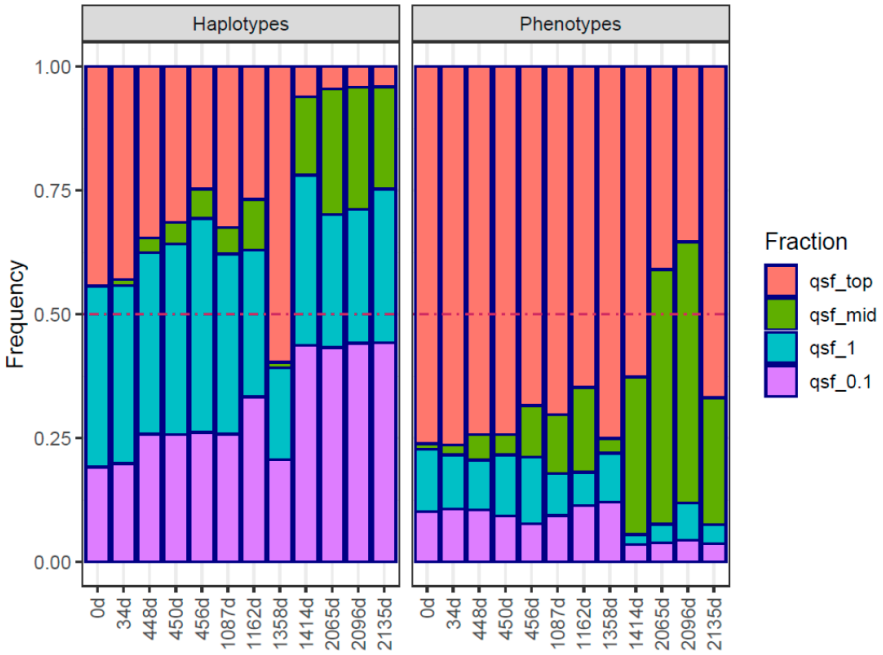


Figure 2. Quasispecies Fitness Fractions for each sample, at the genetic level on the left, and at the phenotypic level on the right. qsf_top: fraction of reads belonging to the master; qsf_mid: fraction of reads belonging to emerging haplotypes; qsf_1: fraction of reads of rare haplotypes; qsf_0.1: fraction of reads of very rare and defective haplotypes. The dash-dot line at 0.5 helps in visualizing the masters of either type with a frequency above 50%.

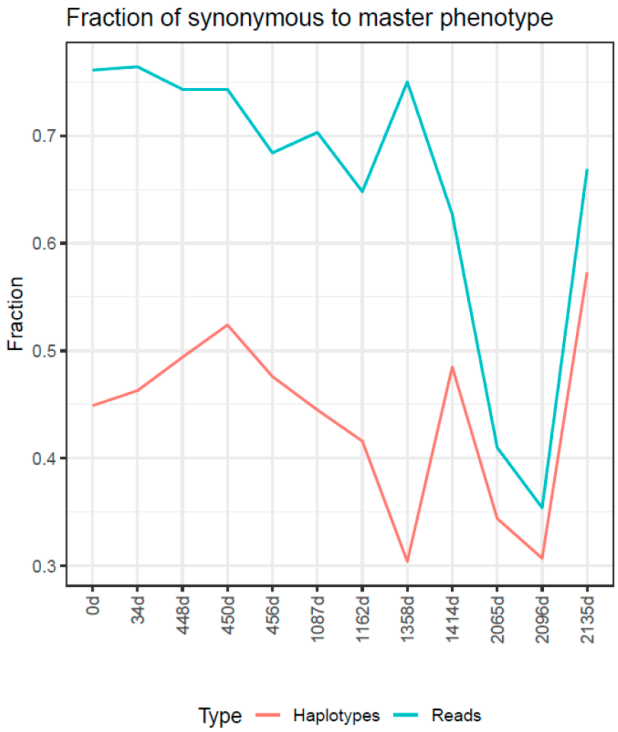


Figure 3. The fraction of synonymous reads to the master phenotype of each sample (turquoise) and the fraction of synonymous haplotypes to the master phenotype of each sample (orange).

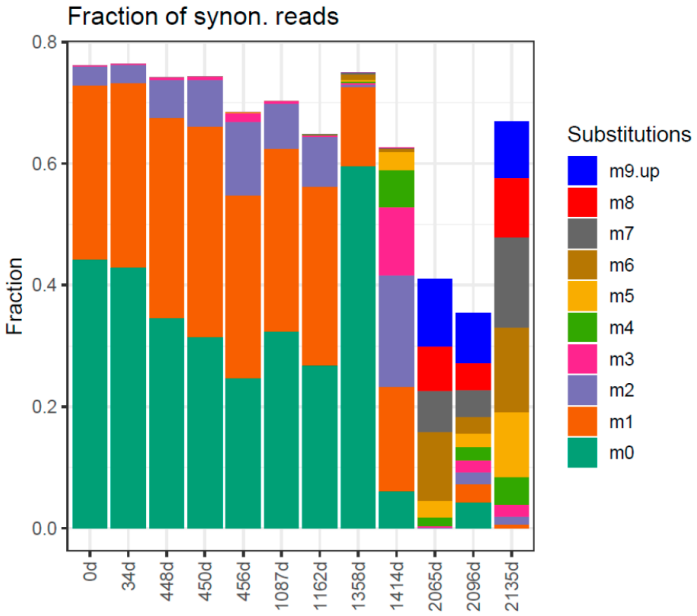


Figure 4. Aggregation of reads of synonymous haplotypes to the master phenotype, according to the number of substitutions with respect to the master haplotype in each sample. m0: fraction of reads for the master; m1: fraction of reads with one substitution with respect to the master haplotype; and so on.

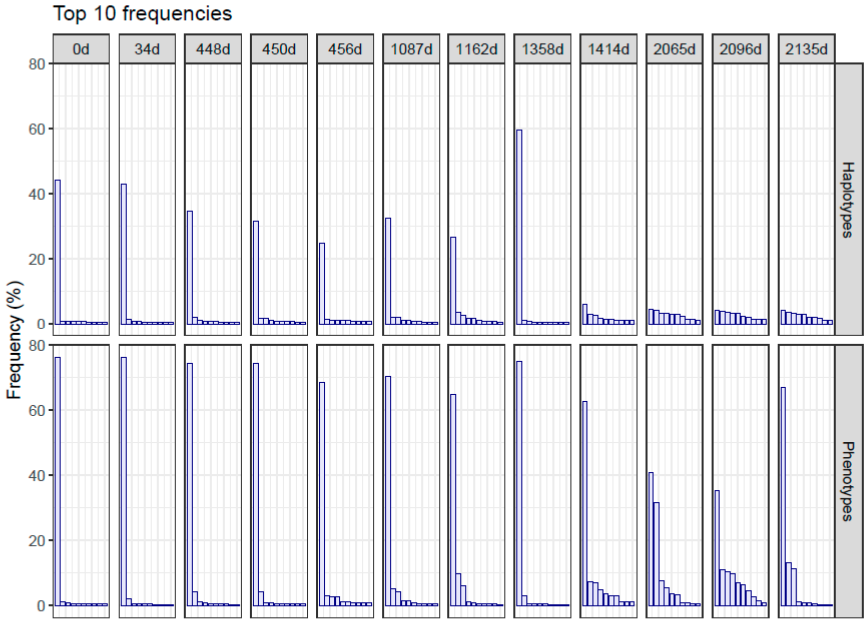


Figure 5. Frequencies of the top 10 haplotypes and phenotypes, shown in decreasing order.

Traits in Quasispecies Evolution

The first treatment lasted from day 447 to 536, with samples at 448 (S03), 450 (S04), and 456d (S05). Viral load dropped slightly from 4.00×10^6 to 2.78×10^6 and 1.17×10^6. The master haplotype decreased in frequency (34.58%, 31.49%, and 24.72%). The quasispecies showed steady fraction <0.1% in QFF, increasing in 0.1–1% and emergent (>1%). Similar patterns emerged at the phenotypic level (Figure 2). RBV's effect could be seen in declining similarity between 450d/448d and 456d/450d (Supplementary Figure S8) and the increase in intra-quasispecies diversity (Supplementary Figure S10).

After this EOT and during the absence of RBV treatment, the same master haplotype and phenotype were maintained between samples 1087d (S06) and 1162d (S07), showing similar frequencies (32.47%, 26,81% and 70.30%, 64.77%) as before (Figure 1). The quasispecies structure and diversity remained similar in terms of QFF (Figure 2) and inter-sample similarity and distance (Supplementary Figures S8 and S9). Amino acid diversity increased, whereas nucleotidic diversity remained steady (Supplementary Figure S10), and the fraction of synonymous reads decreased (Figure 3).

On day 1162 (S07), the patient entered a clinical trial with RBV 200/400 mg daily. No samples were taken in this 20-week period. After the second EOT (no RNA-VHE negativization), at day 1358 (S08), the master haplotype and phenotype frequencies increased significantly (59.67% and 75.05%, Figure 1). Quasispecies diverge from the previous sample (1162 (S07), previous to the trial) both genetically and phenotypically (Supplementary Figures S4, S8 and S9). The QFF showed a tiny fraction of emergent haplotypes and reduced rare haplotypes (Figure 2). The fraction of synonymous reads expressing the master phenotype increased significantly, from 64.7% to 75.0%, whereas the fraction of synonymous haplotypes decreased from 41.6% to 30.4% (Figure 3), suggesting that a new quasispecies might have appeared as a consequence of the proliferation of new fitness-enhanced haplotypes and phenotypes caused by the mutagenic treatment.

After 56 days of no treatment, in 1414d (S09), a different master haplotype and phenotype emerged (6.1% and 62.7%). Genetic QFF showed proliferation of haplotypes, especially emergent ones and the 0.1 fitness fraction, whereas only emerging phenotypes increased at this time point (Figure 2). Interestingly, the similarity with the previous sample was

low for haplotypes but relatively high for phenotypes (functional level). Both nucleotide and amino acid diversity (Supplementary Figure S10) increased substantially due to the proliferation of new genomes progressing from very low frequencies.

Finally, after 80 weeks, a new treatment of 400 mg daily for 24 weeks is started. Three samples at 2065 (S13), 2096 (S15), and 2135 days (S18) showed different master haplotypes at low frequencies (4.58%, 4.24%, and 4.12%) but the same phenotype (40.96%, 35.42%, and 66.86%) (Figure 1). Quasispecies clustered with the 1414d (S09) sample but showed differences (Supplementary Figure S4). Genetic QFF resembled the 1414d structure, but phenotypic QFF indicates emerging phenotypes at the expense of the master, except for the last sample (Figure 2). The similarity in sequential sample haplotypes remained low, whereas phenotypes showed values around 0.5 (Supplementary Figure S8). Viral loads stayed above or slightly below 5 logs, indicating a poor treatment response.

In summary, the QFF at the genetic level shows a complex quasispecies, with multiple haplotypes able to compete with the master, and eventually replace it, with a steady and very important fraction of reads for haplotypes <1% in frequency, which remain at the same level despite the time between them. The QFF at the phenotypic level shows also an important fraction of emerging phenotypes, which represent functional alternatives to the master. To further clarify this situation and the mutagenic effects of the treatment, the reads of haplotypes synonymous to the master phenotype in each sample were aggregated according to the number of substitutions in the haplotype with respect to the master haplotype in each sample. The result is shown in Figure 4. All samples until 1162d (S07) show a high fraction of single mutant haplotypes (m1), similar to the frequency of the master haplotype (m0), and a growing fraction of double mutant haplotypes (m2). At 1358d (S08), the master and the single mutants are the only important fractions, but tiny fractions of higher order mutants are observed. These higher order mutants were undoubtedly generated during the treatment in the clinical trial and could proliferate from very low frequencies to appear visible in the 8 weeks from EOT. The 1414d sample (S09), 8 weeks later, confirmed this proliferation with sensible fractions at multiplicities 1 to 5 (Figure 4). The last three samples (2065d, 2096d, and 2135d), again under treatment but with relatively high viral loads, showed the further proliferation of old mutants and newly generated with multiplicities up to 15 substitutions, observed in this ORF2 amplicon of 363 bp.

Regarding the pattern of substitutions with respect to the consensus one, we observed that the majority of substitutions were transitions. Interestingly, substitutions such as C $\rightarrow$ T/C and G $\rightarrow$ A/G, associated with the ribavirin effect, were notably prevalent in the last four samples (Supplementary Table S6). This pattern is accentuated when specifically analyzing synonymous nucleotide mutations (Figure 6).

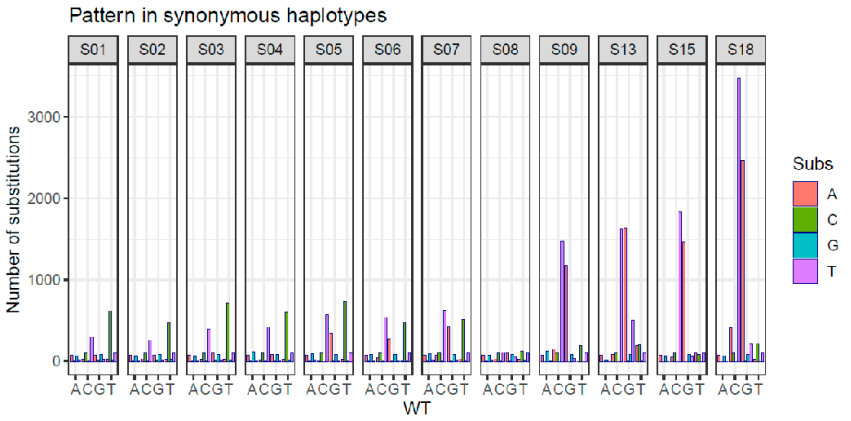


Figure 6. Substitution pattern in synonymous haplotypes to the wild type (consensus synonymous haplotype in each sample).

Beyond the synonymous haplotypes, of whatever multiplicity, generated during the treatment, the conservative substitutions expressing alternative fitness-enhanced phenotypes should be considered. The distribution of the 10 top haplotypes and phenotypes in each sample are represented, in descending order of frequency, in Figure 5. The last four samples show the top ten haplotypes at low and very similar frequencies, indicating similar fitness. The corresponding phenotype distributions show a prominent master but a series of alternative phenotypes at sensible frequencies. Analyzing the amino acid changes that these haplotypes introduce with respect to the master phenotype in the d0 sample, we see that all could be the product of a single nucleotide substitution, and the resulting amino acids have similar chemical characteristics to the wild type (Table 2 and Supplementary Figure S11) compatible with moderate-to-fitness-enhanced alternative phenotypes.

Table 2. Variants observed in emerging haplotypes of the last four samples with respect to the d0 phenotype. WT: wild type amino acid (in d0 phenotype). Var: observed mutation. N: number of phenotypes in which this mutation has been observed, this includes any master or emerging phenotypes of any of the four samples. Fitch: distance of Fitch as the minimum number of nucleotide substitutions required for the mutation. Grantham: chemical distance between the two amino acids. qGrantham: quantile of the corresponding Grantham distance.

WT	Var	N	Fitch	Grantham	qGrantham
T	N	27	1	65	0.2579
T	A	20	1	58	0.2105
G	S	3	1	56	0.2000
V	I	3	1	29	0.0789
A	T	2	1	58	0.2105
E	K	2	1	56	0.2000
I	M	2	1	10	0.0105
L	F	2	1	22	0.0368
A	V	1	1	64	0.2421
F	Y	1	1	22	0.0368
I	V	1	1	29	0.0789
S	A	1	1	99	0.5368
S	P	1	1	74	0.2947
V	A	1	1	64	0.2421
Y	H	1	1	83	0.3421

3. Discussion

In our study, we observed the cumulative effect of three different ribavirin treatment regimens (currently, the only clinical option), with EOTs, and large periods with no treatment in between, and we compared the effect at the nucleotide (genotype) and amino acid (phenotype) level. The master haplotype at mid frequencies (24.7–44.3%) was maintained up to 1162d; a new master haplotype appeared at 1358d with higher frequency (59.7%); and a new master haplotype was generated in each successive sample, with very low frequencies (4.1–6.1%), showing an unstructured collection of genomes when the viral population showed insensitivity to increase in RBV amounts (1414d to 2135d) at the nucleotide level. Interestingly, first treatment at the naïve-treatment time, with 200 mg RBV, seemed to be effective since RNA achieved negativization 90 days after starting treatment, at least with the diagnostic sensitivity used at that time using plasma sample, suggesting that lethality could have been achieved. However, by stopping treatment, possibly too early in time, the virus had not reached complete extinction. Basal viral replication likely generated a drifting mutant cloud with continuous loss of the master sequence due to the advantage of genomes lying on a high fitness plateau [25,26] (reviewed in Tejero et al. [27]) and helped the population to move through the sequence space changing the master sequence in 1358d (frequency of 59.7%) and a subsequent jump to a new master sequence. The most breaking result in our study is that, at the nucleotide level, dynamics of viral quasispecies look like there is a loss in the fitness of the master sequence, but, at the phenotype level, once the

virus found the new master sequence at day 1414d, it becomes highly dominant (66.86% at day 2135d) despite the increase in the amount of RBV provided. Quasispecies theory suggests that viruses could also achieve resistance by moving to flatter regions of the fitness landscape, where the density of neutral mutations is higher (survival of the flattest) [28,29].

Treating chronic HEV infection, primarily affecting immunocompromised patients, is challenging due to the absence of direct-acting antiviral drugs for HEV. The virus's high mutation rate, and high viral loads in these patients pose an added problem, reducing the effectiveness of repurposed drugs like sofosbuvir for antiviral treatment. Currently, Ribavirin (RBV) is the only effective drug against HEV due to its multiple mechanisms of action [30], including a putative lethal mutagenesis effect [17,31].

Lethal mutagenesis is based on the idea that RNA viruses replicate at an exceptional high mutation rate, nearing what's known as error catastrophe [32,33]. Eigen's original theory proposes that a quasispecies can remain stable despite a high mutation rate. However, even small increases in mutation rate can disrupt this equilibrium, leading to loss of the master sequence and meaningful genetic information due to a surge in errors [34].

Initially, error catastrophe described the deterioration of cellular functions in the context of aging [35]. It was later adapted to the quasispecies theory [36], signifying a critical transition from a structured distribution of viral genomes, usually dominated by a master sequence, to an unstructured collection of genomes, without a master sequence. Lethal mutagenesis aims to render a viral population nonfunctionally by elevating the average mutation rate beyond a specific threshold, often achieved through the use of mutagenic agents.

RBV-mediated lethal mutagenesis (extinction through an increase in mutant spectrum complexity and decrease in specific infectivity) has been reported for HCV in cell cultures [37] and in other viruses [17]. Early studies on foot-and-mouth disease virus (FMDV) revealed that viral populations with high fitness and large population sizes exhibited reduced susceptibility to mutagenic agents, leading to delayed virus extinction [38]. In our study, the sustained viral loads along the treatments, with low effectivity, could be, in part, explained by the presence of a highly fit viral population.

Previous studies have demonstrated that achieving extinction in high-fitness viral populations requires a strategic combination of mutagenic agents and nonmutagenic inhibitors [39,40]. This approach was found necessary against the resilient nature of such quasispecies. Actually, a higher suppressive effect of a sequential administration of an antiviral non-mutagenic inhibitor, followed by a mutagenic agent than the converse (first a mutagenic agent and then an inhibitor) or the corresponding combination (inhibitor and mutagen-administered together), has been documented [41,42]. The rationale is that the administration of the inhibitor will produce a decrease in viral load, which will render the system more susceptible to mutagenesis-mediated extinction, allowing expression of interfering activities associated with the mutagenized spectrum of mutants [43].

The results in this study are consistent with the development of resistance to the treatment. Whether specifically to RBV, the result of a highly fit and complex quasispecies, or both at the same time, is unknown. The presence of variants over the full HEV polymerase, with these samples, is under study. On the other hand, viral fitness has been described as a key factor in resistance to treatments in HCV, both mutagenic and direct acting agents [16,44].

A similar case of an HEV patient treated with different regimens of RBV has been recently described [16] with the same consequences: high final viral loads of a non-responding quasispecies, highly diverse in genetic and phenotypic terms. Furthermore, these results could be theoretically anticipated. Only two factors are required to achieve this final situation: (a) an enhanced production of mutants (mutagenic treatment, like RBV) and (b) a replicating system (quasispecies in a host) at a pace (viral load) able to keep selecting the most-fit variants while removing the weakest. On the other hand, each treatment discontinuation with RNA+ will bring to the proliferation of the new fitted variants produced

during the treatment, resulting in a highly diverse quasispecies resilient to new treatments. This study shows a realization of this scenario: quasispecies dynamics in pure state.

4. Materials and Methods

4.1. Patient Data

Sequential plasma samples were collected at different time points (Table 1, Supplementary Figure S1) from a 62-year-old patient chronically infected by HEV. His medical records included lung transplantation for interstitial lung disease and sequential kidney transplantation due to end-stage chronic kidney disease. Diagnosis of chronic hepatitis E was established based on detectable HEV RNA after long-term transaminases increased. The first RBV treatment was adjusted to kidney function to 200 mg/day for three months (12 weeks), resulting in an undetectable HEV RNA. One year later, there was a relapse, leading to a second RBV treatment of alternating 200 and 200/400 mg/day for 6 months (24 weeks), with a posterior 12 months (48 weeks) monitorization. The second treatment was not capable of eliminating the HEV but achieved hepatic biochemistry normalization and platelet count elevation. Finally, a third treatment of 400 mg/day for 6 months was applied due to hepatic decompensation and ascites appearance. This third regime was able to reduce ascites but could not eliminate HEV infection nor SVR, keeping a sustained viremia.

4.2. RNA Extraction and Amplification

RNA was extracted using the QIAamp RNA Viral MiniKit (QIAGEN, Hilden, Germany) following the provider protocol. RNA was retrotranscribed and subsequently amplified using SuperScript™ III One-Step RT-PCR System with Platinum™ Taq High Fidelity DNA Polymerase (Termo Fisher Scientific, Carlsbad, CA, USA), and two external ORF2 primers: FW 5′CCGACAGAATTGATTTCGTCGGC3′ and RV 5′ACTCCCGRGTYTTACCYA CCTT3′. RT-PCR was performed at 50 °C for 30 min, followed by 7 min a 94 °C for RT enzyme inactivation, and successive cycling of 94 °C 10 s denaturation, 54 °C 30 s annealing, and 68 °C 1 min 30 s elongation, with a final elongation of 7 min. Nested PCR was carried out with FastStart High Fidelity PCR System, dNTPack (Roche, Basel, Switzerland), according to manufacturer indications using the internal primers FW 5′ GTCGTCTCAGCCAATGGCGAGCC3′ and RV 5′CASARAANGTCTTNGARTACTGCT3′, with annealing and elongation temperature at 50 °C and 72 °C, respectively (Figure 7). PCR product was purified using KAPA Pure Beads (Kapa Biosystems, Roche, Pleasanton, CA, USA). Quantification was carried out by Qubit fluorometry using a double-stranded DNA High Sensitivity Assay Kit (Termo Fisher Scientific, Carlsbad, CA, USA).

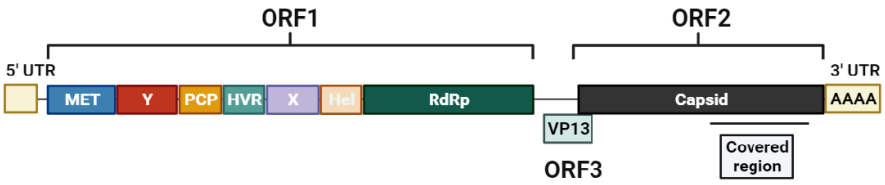


Figure 7. HEV genomic structure scheme pointing out the covered region. Created with BioRender.com.

4.3. Library Preparation and Illumina Sequencing

Library preparation was carried out using KAPA HyperPrep Kit (Kapa Biosystems, Roche, Pleasanton, CA, USA), according to manufacturer instructions. Samples were single indexed using the SeqCap Adapter Kit A/B (Nimblegen, Roche, Pleasanton, CA, USA), and DNA purification was performed through KAPA Pure Beads (Kapa Biosystems, Roche, Pleasanton, CA, USA). Final library concentration was assessed by qPCR using the KAPA Library Quantification Kit (KapaBiosystems, Roche, Pleasanton, CA, USA). Library was loaded onto Miseq Illumina platform through MiSeq Reagent Kit v3 600 cycles (Illumina, San Diego, CA, USA).

4.4. Processing the Sequencing Data

The aim of the sequencing data treatment is to discard error-bearing reads while preserving full-length read integrity so that haplotypes that completely cover the amplicon with their respective frequencies were incorporated. The steps in this process have been previously described [16,45] and may be summarized as follows:

- Obtain Fastq files with Illumina 2 × 300-bp paired-end reads;
- Recover full amplicon reads with FLASH [46] (minimum 20-bp overlap, maximum 10% mismatches). The 300-bp reads, when overlapped, result in reads covering complete ~400–500 bp amplicons;
- Remove full reads with 5% or more bases below a Phred score of Q30;
- Demultiplex and trim primers (max three differences accepted);
- Collapse reads (molecules) to haplotypes (amplicon-genomes) and their frequencies (read counts);
- Multiple alignment of all haplotypes in each sample/amplicon;
- Remove all haplotypes that are not common to both DNA strands and supported at least by 5 reads;
- Remove insertions in master haplotypes and repair single gaps in remaining haplotypes. Recollapse to haplotypes and vectors of frequencies;
- The haplotypes in each sample/amplicon are translated to phenotypes and recollapsed to obtain the set of phenotypes with corresponding frequencies (read counts).

4.5. Bioinformatic Procedures

The haplotypes/phenotypes and corresponding frequencies resulting from the sequencing data are the basis of subsequent computations. A detailed description of these methods is provided in the supplementary material. Briefly, the computations took into account sample size dependence of diversity indices by repeated resampling to the reference coverage (147,000 reads). The quasispecies structure and diversity was studied by the method of the Quasispecies Fitness Fractions and the Hill Numbers Profile. Distances between quasispecies were computed by the Matoshi-Nei method, using raw genetic distances between pairs of haplotypes at the genetic level, or the Grishin distances based on BLOSUM80 matrix values at the phenotypic level.

4.6. Software and Statistics

All computations were performed in R (v4.2.2) [47] with in-house scripts, using the Biostrings [48], ShortRead [49], and QSutils [50] packages from Bioconductor [51], as well as ape [52], bios2mds [53], tidyverse [54], and ggplot2 [55].

5. Conclusions

The treatment with mutagenic agents concomitant with high viral loads resulted in an accelerated evolution of the quasispecies with the selection of fitness-enhanced haplotypes and phenotypes, creating fitness-enhanced and extremely diverse quasispecies that are resistant to further treatments. The prescription of mutagens should only be advised when combined with efficient viral inhibitors, in order to keep the replication as low as possible and to avoid the selection of better-fit variants. Even in this case, it should be prescribed only to patients able to follow the treatment until RNA negativization, with no adverse effects that could recommend an early discontinuation of the treatment. The risk is to end up with a better adapted quasispecies, resilient to further treatments.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/ijms242417185/s1. References [56–61] are cited in the supplementary materials.

Author Contributions: S.C.-C.: Methodology, investigation, visualization, and writing—original draft; J.G.: Conceptualization, methodology, software, formal data analysis, visualization, and writing—original draft; D.G.-C.: Methodology, investigation, and project administration; M.R.-B.: Resources, data analysis, and validation; M.B.: Conceptualization, resources, and writing—review and editing; A.R.-S.: investigation and resources; J.V.-R.: investigation and resources; C.C.: Investigation and visualization; M.I.-L.: Software and data curation; C.M.A.: Investigation and visualization; M.F.C.: Methodology and visualization; D.T.: Methodology and visualization; J.I.E.: Conceptualization, resources, validation, and writing—review and editing; F.R.-F.: Conceptualization, funding acquisition, data analysis, and writing—review and editing; J.Q.: Conceptualization, methodology, validation, supervision, funding acquisition, and writing—original draft. All authors have read and agreed to the published version of the manuscript.

Funding: This study was partially supported by Plan Estratègic de Recerca i Innovació en Salut (PERIS)—Direcció General de Recerca i Innovació en Salut (DGRIS), Catalan Health Ministry, Generalitat de Catalunya; Centro para el Desarrollo Tecnológico Industrial (CDTI) from the Spanish Ministry of Economy and Business, grant number IDI-20200297, Grant PID2021-126447OB-I00 funded by MCIN/AEI/10.13039/501100011033 and by ERDF A way of making Europe; Projects PI19/00301 and PI22/00258 funded by Instituto de Salud Carlos III (ISCIII) and cofounded by the European Union; and Gilead's biomedical research project GLD21/00006. S.C.-C has received support from Spanish Ministry of Education, grant FPU21/04150. M.I.-L. received the support of a fellowship from the "la Caixa" Foundation (ID 100010434), whose code is "LCF/BQ/DR23/12000020".

Institutional Review Board Statement: The study was approved by Vall d'Hebron University Hospital Ethics Committee, with reference number PR(AG)287-2022.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The genomic nucleotide sequences included in this study are being submitted in the GENBank repository database as Bioproject ID PRJNA1038697.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. World Health Organization (WHO). Hepatitis E. Available online: https://www.who.int/news-room/fact-sheets/detail/hepatitis-e (accessed on 29 November 2023).
2. Ministerio de Sanidad. 2º Estudio de Seroprevalencia En España. Available online: https://www.sanidad.gob.es/profesionales/saludPublica/prevPromocion/vacunaciones/comoTrabajamos/docs/EstudioSeroprevalencia_EnfermedadesInmunoprevenibles.pdf (accessed on 29 November 2023).
3. Sauleda, S.; Ong, E.; Bes, M.; Janssen, A.; Cory, R.; Babizki, M.; Shin, T.; Lindquist, A.; Hoang, A.; Vang, L.; et al. Seroprevalence of Hepatitis E Virus (HEV) and Detection of HEV RNA with a Transcription-Mediated Amplification Assay in Blood Donors from Catalonia (Spain). *Transfusion* **2015**, *55*, 972–979. [CrossRef] [PubMed]
4. Purdy, M.A.; Harrison, T.J.; Jameel, S.; Meng, X.-J.; Okamoto, H.; Van der Poel, W.H.M.; Smith, D.B. ICTV Report Consortium ICTV Virus Taxonomy Profile: Hepeviridae. *J. Gen. Virol.* **2017**, *98*, 2645–2646. [CrossRef] [PubMed]
5. Kamar, N.; Dalton, H.R.; Abravanel, F.; Izopet, J. Hepatitis E Virus Infection. *Clin. Microbiol. Rev.* **2014**, *27*, 116–138. [CrossRef]
6. Riveiro-Barciela, M.; Minguez, B.; Girones, R.; Rodriguez-Frias, F.; Quer, J.; Buti, M. Phylogenetic Demonstration of Hepatitis E Infection Transmitted by Pork Meat Ingestion. *J. Clin. Gastroenterol.* **2015**, *49*, 165–168. [CrossRef] [PubMed]
7. Boxman, I.L.A.; Jansen, C.C.C.; Hägele, G.; Zwartkruis-Nahuis, A.; Tijsma, A.S.L.; Vennema, H. Monitoring of Pork Liver and Meat Products on the Dutch Market for the Presence of HEV RNA. *Int. J. Food Microbiol.* **2019**, *296*, 58–64. [CrossRef]
8. Hoofnagle, J.H.; Nelson, K.E.; Purcell, R.H. Hepatitis E. *N. Engl. J. Med.* **2012**, *367*, 1237–1244. [CrossRef]
9. Dalton, H.R.; Kamar, N.; Baylis, S.A.; Moradpour, D.; Wedemeyer, H.; Negro, F.; European Association for the Study of the Liver. EASL Clinical Practice Guidelines on Hepatitis E Virus Infection. *J. Hepatol.* **2018**, *68*, 1256–1271. [CrossRef]
10. Bernal, W.; Auzinger, G.; Dhawan, A.; Wendon, J. Acute Liver Failure. *Lancet* **2010**, *376*, 190–201. [CrossRef]
11. Zhou, X.; de Man, R.A.; de Knegt, R.J.; Metselaar, H.J.; Peppelenbosch, M.P.; Pan, Q. Epidemiology and Management of Chronic Hepatitis E Infection in Solid Organ Transplantation: A Comprehensive Literature Review. *Rev. Med. Virol.* **2013**, *23*, 295–304. [CrossRef]
12. Kamar, N.; Izopet, J.; Pavio, N.; Aggarwal, R.; Labrique, A.; Wedemeyer, H.; Dalton, H.R. Hepatitis E Virus Infection. *Nat. Rev. Dis. Primers* **2017**, *3*, 17086. [CrossRef]
13. Rivero-Juarez, A.; Vallejo, N.; Lopez-Lopez, P.; Díaz-Mareque, A.I.; Frias, M.; Vallejo, A.; Caballero-Gómez, J.; Rodríguez-Velasco, M.; Molina, E.; Aguilera, A. Ribavirin as a First Treatment Approach for Hepatitis E Virus Infection in Transplant Recipient Patients. *Microorganisms* **2019**, *8*, 51. [CrossRef] [PubMed]

14. Kamar, N.; Izopet, J.; Tripon, S.; Bismuth, M.; Hillaire, S.; Dumortier, J.; Radenne, S.; Coilly, A.; Garrigue, V.; D'Alteroche, L.; et al. Ribavirin for Chronic Hepatitis E Virus Infection in Transplant Recipients. *N. Engl. J. Med.* **2014**, *370*, 1111–1120. [CrossRef] [PubMed]
15. Debing, Y.; Emerson, S.U.; Wang, Y.; Pan, Q.; Balzarini, J.; Dallmeier, K.; Neyts, J. Ribavirin Inhibits in Vitro Hepatitis E Virus Replication through Depletion of Cellular GTP Pools and Is Moderately Synergistic with Alpha Interferon. *Antimicrob. Agents Chemother.* **2014**, *58*, 267–273. [CrossRef]
16. Gregori, J.; Colomer-Castell, S.; Campos, C.; Ibañez-Lligoña, M.; Garcia-Cehic, D.; Rando-Segura, A.; Adombi, C.M.; Pintó, R.; Guix, S.; Bosch, A.; et al. Quasispecies Fitness Partition to Characterize the Molecular Status of a Viral Population. Negative Effect of Early Ribavirin Discontinuation in a Chronically Infected HEV Patient. *Int. J. Mol. Sci.* **2022**, *23*, 14654. [CrossRef]
17. Perales, C.; Gallego, I.; de Ávila, A.I.; Soria, M.E.; Gregori, J.; Quer, J.; Domingo, E. The Increasing Impact of Lethal Mutagenesis of Viruses. *Future Med. Chem.* **2019**, *11*, 1645–1657. [CrossRef] [PubMed]
18. Hassine, I.H.; Ben M'hadheb, M.; Menéndez-Arias, L. Lethal Mutagenesis of RNA Viruses and Approved Drugs with Antiviral Mutagenic Activity. *Viruses* **2022**, *14*, 841. [CrossRef]
19. Schulz, M.; Papp, C.P.; Bock, C.-T.; Hofmann, J.; Gerlach, U.A.; Maurer, M.M.; Eurich, D.; Mueller, T. Combination Therapy of Sofosbuvir and Ribavirin Fails to Clear Chronic Hepatitis E Infection in a Multivisceral Transplanted Patient. *J. Hepatol.* **2019**, *71*, 225–227. [CrossRef]
20. Sastre, L.; García-López, M.; Pérez-Del-Pulgar, S.; Lens, S.; Costa, J.; Navasa, M.; Forns, X. The Challenge of Chronic Hepatitis E in Liver Transplant Recipients: Failure of Sofosbuvir plus Ribavirin Therapy. *Gastroenterol. Hepatol.* **2020**, *43*, 136–137. [CrossRef]
21. Takahashi, K.; Toyota, J.; Karino, Y.; Kang, J.H.; Maekubo, H.; Abe, N.; Mishiro, S. Estimation of the Mutation Rate of Hepatitis E Virus Based on a Set of Closely Related 7.5-Year-Apart Isolates from Sapporo, Japan. *Hepatol. Res.* **2004**, *29*, 212–215. [CrossRef]
22. Holland, J.J.; Spindler, K.; Horodyski, F.; Grabau, E.; Nichol, S.; VandePol, S. Rapid Evolution of RNA Genomes. *Science* **1982**, *215*, 1577–1585. [CrossRef]
23. Vignuzzi, M.; Stone, J.K.; Arnold, J.J.; Cameron, C.E.; Andino, R. Quasispecies Diversity Determines Pathogenesis through Cooperative Interactions in a Viral Population. *Nature* **2006**, *439*, 344–348. [CrossRef]
24. Domingo, E. Virus as Populations. Composition, Complexity, Dynamics and Biological Implications. In *Virus as Populations*, 1st ed.; Academic Press: Cambridge, MA, USA; Elsevier: London, UK, 2016; pp. 1–412.
25. Wilke, C.O.; Wang, J.L.; Ofria, C.; Lenski, R.E.; Adami, C. Evolution of Digital Organisms at High Mutation Rates Leads to Survival of the Flattest. *Nature* **2001**, *412*, 331–333. [CrossRef] [PubMed]
26. Schuster, P.; Swetina, J. Stationary Mutant Distributions and Evolutionary Optimization. *Bull. Math. Biol.* **1988**, *50*, 635–660. [CrossRef] [PubMed]
27. Tejero, H.; Montero, F.; Nuño, J.C. Theories of Lethal Mutagenesis: From Error Catastrophe to Lethal Defection. *Curr. Top. Microbiol. Immunol.* **2016**, *392*, 161–179. [CrossRef] [PubMed]
28. Wilke, C.O. Quasispecies Theory in the Context of Population Genetics. *BMC Evol. Biol.* **2005**, *5*, 44. [CrossRef]
29. Lauring, A.S.; Andino, R. Quasispecies Theory and the Behavior of RNA Viruses. *PLoS Pathog.* **2010**, *6*, e1001005. [CrossRef]
30. Beaucourt, S.; Vignuzzi, M. Ribavirin: A Drug Active against Many Viruses with Multiple Effects on Virus Replication and Propagation. Molecular Basis of Ribavirin Resistance. *Curr. Opin. Virol.* **2014**, *8*, 10–15. [CrossRef]
31. De Clercq, E.; Li, G. Approved Antiviral Drugs over the Past 50 Years. *Clin. Microbiol. Rev.* **2016**, *29*, 695–747. [CrossRef]
32. Schuster, P. Quasispecies on Fitness Landscapes. *Curr. Top. Microbiol. Immunol.* **2016**, *392*, 61–120. [CrossRef]
33. Swetina, J.; Schuster, P. Self-Replication with Errors. A Model for Polynucleotide Replication. *Biophys. Chem.* **1982**, *16*, 329–345. [CrossRef]
34. Eigen, M.; Schuster, P. The Hypercycle. A Principle of Natural Self-Organization. Part A: Emergence of the Hypercycle. *Naturwissenschaften Sci. Nat.* **1977**, *64*, 541–565. [CrossRef]
35. Orgel, L.E. The Maintenance of the Accuracy of Protein Synthesis and Its Relevance to Ageing. *Proc. Natl. Acad. Sci. USA* **1963**, *49*, 517–521. [CrossRef]
36. Eigen, M.; Schuster, P. *The Hypercycle. A Principle of Natural Self-Organization*; Springer: Berlin/Heidelberg, Germany, 1979.
37. Ortega-Prieto, A.M.; Sheldon, J.; Grande-Perez, A.; Tejero, H.; Gregori, J.; Quer, J.; Esteban, J.I.; Domingo, E.; Perales, C. Extinction of Hepatitis C Virus by Ribavirin in Hepatoma Cells Involves Lethal Mutagenesis. *PLoS ONE* **2013**, *8*, e71039. [CrossRef] [PubMed]
38. Sierra, S.; Dávila, M.; Lowenstein, P.R.; Domingo, E. Response of Foot-and-Mouth Disease Virus to Increased Mutagenesis: Influence of Viral Load and Fitness in Loss of Infectivity. *J. Virol.* **2000**, *74*, 8316–8323. [CrossRef] [PubMed]
39. Tapia, N.; Fernàndez, G.; Parera, M.; Gómez-Mariano, G.; Clotet, B.; Quiñones-Mateu, M.; Domingo, E.; Martínez, M.A. Combination of a Mutagenic Agent with a Reverse Transcriptase Inhibitor Results in Systematic Inhibition of HIV-1 Infection. *Virology* **2005**, *338*, 1–8. [CrossRef] [PubMed]
40. Pariente, N.; Sierra, S.; Lowenstein, P.R.; Domingo, E. Efficient Virus Extinction by Combinations of a Mutagen and Antiviral Inhibitors. *J. Virol.* **2001**, *75*, 9723–9730. [CrossRef] [PubMed]
41. Perales, C.; Agudo, R.; Tejero, H.; Manrubia, S.C.; Domingo, E. Potential Benefits of Sequential Inhibitor-Mutagen Treatments of RNA Virus Infections. *PLoS Pathog.* **2009**, *5*, e1000658. [CrossRef] [PubMed]
42. Moreno, H.; Grande-Pérez, A.; Domingo, E.; Martín, V. Arenaviruses and Lethal Mutagenesis. Prospects for New Ribavirin-Based Interventions. *Viruses* **2012**, *4*, 2786–2805. [CrossRef]

43. Huang, A.S.; Baltimore, D. Defective Viral Particles and Viral Disease Processes. *Nature* **1970**, *226*, 325–327. [CrossRef]
44. Gallego, I.; Gregori, J.; Soria, M.E.; Garcia-Crespo, C.; Garcia-Alvarez, M.; Gomez-Gonzalez, A.; Valiergue, R.; Gomez, J.; Esteban, J.I.; Quer, J.; et al. Resistance of High Fitness Hepatitis C Virus to Lethal Mutagenesis. *Virology* **2018**, *523*, 100–109. [CrossRef]
45. Adombi, C.M.; Bosch, A.; Buti, M.; Campos, C.; Colomer-Castel, S.; Cortese, M.F.; Domingo, E.; Esteban, J.I.; Gallego, I.; Garcia-Cehic, D.; et al. *Viral Quasispecies Diversity and Evolution: A Bioinformatics Molecular Approach*; Gregori, J., Rodriguez-Frias, F., Quer, J., Eds.; Il Pensiero Scientifico Editore: Rome, Italy, 2023; ISBN 9788849007589.
46. Magoc, T.; Salzberg, S.L. FLASH: Fast Length Adjustment of Short Reads to Improve Genome Assemblies. *Bioinformatics* **2011**, *27*, 2957–2963. [CrossRef]
47. R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2019.
48. Pages, H.; Aboyoun, P.; Gentleman, R.; DebRoy, S. Biostrings: String Objects Representing Biological Sequences, and Matching Algorithms, R Package 2.38.4. 2023. Available online: https://www.bioconductor.org/packages//2.7/bioc/html/Biostrings.html (accessed on 4 December 2023).
49. Morgan, M.; Anders, S.; Lawrence, M.; Aboyoun, P.; Pages, H.; Gentleman, R. ShortRead: A Bioconductor Package for Input, Quality Assessment and Exploration of High-Throughput Sequence Data. *Bioinformatics* **2009**, *25*, 2607–2608. [CrossRef]
50. Guerrero-Murillo, M.; Gregori, J. QSutils: Quasispecies Diversity. R Package Version 1.0.0. Available online: https://bioconductor.org/packages/release/bioc/html/QSutils.html (accessed on 2 December 2023).
51. Gentleman, R.C.; Carey, V.J.; Bates, D.M.; Bolstad, B.; Dettling, M.; Dudoit, S.; Ellis, B.; Gautier, L.; Ge, Y.; Gentry, J.; et al. Bioconductor: Open Software Development for Computational Biology and Bioinformatics. *Genome Biol.* **2004**, *5*, R80. [CrossRef]
52. Paradis, E.; Schliep, K. Ape 5.0: An Environment for Modern Phylogenetics and Evolutionary Analyses in R. *Bioinformatics* **2019**, *35*, 526–528. [CrossRef]
53. Pelé, J.; Bécu, J.-M.; Abdi, H.; Chabbert, M. Bios2mds: An R Package for Comparing Orthologous Protein Families by Metric Multidimensional Scaling. *BMC Bioinform.* **2012**, *13*, 133. [CrossRef] [PubMed]
54. Wickham, H. Welcome to Master the Tidyverse. *J. Open Source Softw.* **2019**, *4*, 1686. [CrossRef]
55. Valero-Mora, P.M. ggplot2: Elegant graphics for data analysis. *J. Stat. Softw.* **2010**, *35*, 1–3. [CrossRef]
56. Gregori, J.; Soria, M.E.; Gallego, I.; Guerrero-Murillo, M.; Esteban, J.I.; Quer, J.; Perales, C.; Domingo, E. Rare Haplotype Load as Marker for Lethal Mutagenesis. *PLoS ONE* **2018**, *13*. [CrossRef]
57. Nei, M. *Molecular Evolutionary Genetics*; Columbia University Press: New York, NY, USA, 1987; ISBN 9780231886710.
58. Grishin, V.N.; Grishin, N. V Euclidian Space and Grouping of Biological Objects. *Bioinformatics* **2002**, *18*, 1523–1534. [CrossRef]
59. Gregori, J.; Ibañez-Lligoña, M.; Quer, J. Quantifying In-Host Quasispecies Evolution. *Int. J. Mol. Sci.* **2023**, *24*. [CrossRef]
60. Fitch, W.M. An Improved Method of Testing for Evolutionary Homology. *J. Mol. Biol.* **1966**, *16*, 9–16. [CrossRef]
61. Grantham, R. Amino Acid Difference Formula to Help Explain Protein Evolution. *Science* **1974**, *185*, 862–864. [CrossRef]

4.2.3 Article 2: Supplementary Material

In-Host HEV Quasispecies Evolution Shows the Limits of Mutagenic Antiviral Treatments.

Article

In-Host HEV Quasispecies Evolution Shows the Limits of Mutagenic Antiviral Treatments

Sergi Colomer-Castell [1,2,3,†], Josep Gregori [1,*,†], Damir Garcia-Cehic [1,2], Mar Riveiro-Barciela [1,2,4], Maria Buti [1,2,4], Ariadna Rando-Segura [2,5], Judit Vico-Romero [1], Carolina Campos [1,2,3], Marta Ibañez-Lligoña [1,4], Caroline Melanie Adombi [1,6], Maria Francesca Cortese [2,7], David Tabernero [2,7], Juan Ignacio Esteban [1,2,4], Francisco Rodriguez-Frias [2,7] and Josep Quer [1,2,3,*]

[1] Liver Diseases-Viral Hepatitis, Liver Unit, Vall d'Hebron Institut de Recerca (VHIR), Vall d'Hebron Hospital Universitari, Vall d'Hebron Barcelona Hospital Campus, Passeig Vall d'Hebron 119-129, 08035 Barcelona, Spain; sergi.colomer@vhir.org (S.C.-C.); damir.garcia@vhir.org (D.G.-C.); mar.riveiro@gmail.com (M.R.-B.); maria.buti@vallhebron.cat (M.B.); judit.vico@vhir.org (J.V.-R.); carolina.campos@vhir.org (C.C.); marta.ibanez@vhir.org (M.I.-L.); adombicaro@yahoo.fr (C.M.A.); juanignacio.esteban@vallhebron.cat (J.I.E.)

[2] Centro de Investigación Biomédica en Red de Enfermedades Hepáticas y Digestivas (CIBERehd), Instituto de Salud Carlos III, Av. Monforte de Lemos, 3-5, 28029 Madrid, Spain; ariadna.rando@vallhebron.cat (A.R.-S.); maria.cortese@vhir.org (M.F.C.); dtaberc@gmail.com (D.T.); frarodri@gmail.com (F.R.-F.)

[3] Biochemistry and Molecular Biology Department, Universitat Autònoma de Barcelona (UAB), Campus de la UAB, Plaça Cívica, 08193 Bellaterra, Spain

[4] Medicine Department, Universitat Autònoma de Barcelona (UAB), Campus de la UAB, Plaça Cívica, 08193 Bellaterra, Spain

[5] Microbiology Department, Vall d'Hebron Institut de Recerca (VHIR), Vall d'Hebron Hospital Universitari, Vall d'Hebron Barcelona Hospital Campus, Passeig Vall d'Hebron 119-129, 08035 Barcelona, Spain

[6] Institute of Agropastoral Management, University Peleforo Gon Coulibaly, Korhogo BP 1328, Côte d'Ivoire

[7] Biochemistry Department, Vall d'Hebron Institut de Recerca (VHIR), Vall d'Hebron Hospital Universitari, Vall d'Hebron Barcelona Hospital Campus, Passeig Vall d'Hebron 119-129, 08035 Barcelona, Spain

* Correspondence: josep.gregori@gmail.com or josep.gregori@vhir.org (J.G.); josep.quer@vhir.org (J.Q.)

† These authors contributed equally to this work.

SUPPLEMENTARY METHODS

Sample size dependence. Rarefaction.

Diversity indices are dependent to a varying extent on the sample size[1], and this dependence has to be considered when comparing values from different samples. We used rarefaction to correct differences due to sample size. Briefly, from the set of samples to be compared, the minimum coverage (147,463 reads) was taken as the sample size reference. Each sample then underwent 1000 resample cycles, taking the frequencies of all haplotypes in the sample as probabilities, and the reference size as the resampling size. In each cycle, each diversity index was calculated from the resulting sampling. At the end of the 1000 cycles, averages and standard deviations were computed for each diversity index. In this way all diversity index values were normalized to the same sample size (coverage).

Quasispecies fitness partition

At a given time, a quasispecies is usually comprised of a highly predominant (master) haplotype, a few low- to medium-frequency genomes, various rare haplotypes with very low fitness but still able to replicate to some level, and some defective genomes unable to replicate[2]. Within the characteristic dynamics of a quasispecies, the fitness of a haplotype may be approximated by its frequency in the quasispecies. As a consequence, a quasispecies structure, at a given time, may be summarized by aggregating the molecules (reads) that belong to each fitness type, in four fractions, defined as follows:

- Master: the fraction of molecules belonging to the most frequent haplotype; that is, the one present at the highest percentage;
- Emerging: the fraction of molecules present at a frequency >1% and less than the master percentage, belonging to haplotypes that are able to compete with the predominant one and possibly replace it;
- Low fitness: the fraction of molecules present at frequencies from 0.1% to 1%, belonging to haplotypes that have a low probability of progressing to higher frequencies;
- Very low fitness: the fraction of molecules present at frequencies <0.1% belonging to haplotypes with very low fitness and to defective genomes. The likely fate of these molecules individually is degradation, but the fraction is continuously fed with new very low fitness genomes produced by replication errors or by host editing activities.

QFF values were rarefied, as described above.

Distance between quasispecies

A quasispecies can be seen as a genetic population, and the methods used to study diversification in genetic populations can also be used to determine the distance or dissimilarity between different quasispecies. The method of Matoshi Nei[3] was used to compute the net genetic distance between quasispecies, D_A, from raw genetic distances between pairs of haplotypes in both quasispecies. The net distance, D_A, is the average distance between pairs of haplotypes or phenotypes in the two quasispecies, D_{XY}, corrected by the mean of intra-quasispecies genetic or phenotypic diversity of the two samples being compared.

The same method was used to compute the distance between quasispecies in terms of phenotypic distances, where the distance between pairs of phenotypes was computed by the method of Grishin[4] applied to BLOSUM80 matrix values. D_A values have been divided by the maximum in the series to obtain values in the range 0-1, and render comparable the evolution at the genetic and at the phenotypic level.

On the other hand, the similarity between pairs of sequential quasispecies was computed by comparing the haplotype/phenotype quasispecies distributions using the indices of common molecules (*Cm*), distribution overlap (*Ov*), and Yue-Clayton (*YC*)[5]. These similarities, normalized in the range 0-1, are transformed into dissimilarities by the rule: *Dissimilarity = 1 – similarity*.

From any of these distances or dissimilarities, quasispecies dendrograms and MDS maps are constructed to visualize the relationship between them.

Given that all haplotypes or phenotypes in all samples to be compared must be multiple aligned before computing distances between pairs of them, and the high number of haplotypes in the present study, an approximation was used. The top 50 (most abundant) haplotypes, or the top 20 phenotypes, of each sample were selected to be aligned and used in the computations. As the terms in the computation of distances are of order 2, their value quickly fade for very low frequencies, and the approximation is valid in the context of this study (see Figure 5). The impact of the approximation in terms of reads is also shown in Supplementary. Figure S4.

Fraction of synonymous reads and haplotypes

The values represented in Figure 3 have been rarefied to the minimum coverage of 147,463 reads, as described above.

Impact of amino acid changes in protein functionality

The observed amino acid mutations with respect to the master phenotype were evaluated in terms of probability of occurrence by the Fitch distance[6], which states the minimum number of nucleotide substitutions needed to cause the mutation; and in terms of functionality impact by the Grantham distance between amino acid pairs[7]. The Grantham distance between amino acids considers three factors: molecular volume, polarity, and molecular composition. Supplementary Figure S11 shows the combination of the two terms Fitch + Grantham, as a scale where to evaluate the putative impact of a mutation in protein functionality.

SUPPLEMENTARY RESULTS

Coverage and sample characteristics

Supplementary Table S1 summarizes clinical data and sequencing results of all samples in the study. We have studied between 147,463 reads of a single amplicon in ORF2 in sample S09, to 419,804 reads in sample S08. During RBV treatment, we observed a fluctuation in the number of haplotypes and in the frequency of the master sequence.

Master haplotypes vs master phenotypes

The first analysis concerns the confrontation of the dominant sequences at the nucleotide and at the amino acid level in each sequential sample. Figure 1 shows the UPGMA tree of the master haplotypes based on raw nucleotide distances in pairwise-deletion mode, and the corresponding tree with the master phenotypes based in distances computed from the BLOSUM-80 matrix.

At the nucleotide level, the same master haplotype was maintained until past 1162 days post-diagnosis (1162d, S07), although with declining frequencies. The sample at 1414d (S09) shows a different master at relatively low abundance (6.08%). The masters become different at each sequential sample since 1414d (S09). Notably the frequency increases to 59.67% at 1358d (S08), but decreases afterwards, remaining below 5%.

At the functional level (amino acid analysis), the same phenotype is observed until past 1162d (S07) at frequencies between 64.77% and 76.37%. At 1358d (S08) a different phenotype is observed at 75.05%. Finally, a new master phenotype is observed at 1414d (S09) and maintained till the last sample, but declining in frequency (62.3%, 41.0%, 35.4%) until the last sample in which it is expressed at higher frequency (66.9%).

The much larger frequencies of the master phenotypes compared to the master haplotypes, in the samples since 1414d (S09), suggest the presence of a set of haplotypes synonymous to the master phenotype.

As a complement, Supplementary Figure S2 shows the polymorphic sites in the comparison of master haplotypes and phenotypes; and Supplementary Figure S3 shows the track of masters, in frequency, in the sequential samples.

Quasispecies dendrograms

Aligning the top 50 haplotypes in each sample we obtain the matrix of pair-wise quasispecies genetic distances. Based on these distances a UPGMA quasispecies tree is constructed. On the other hand, by the alignment of the to 20 phenotypes in each sample we obtain the pair-wise quasispecies phenotype distances, and the corresponding UPGMA tree is constructed. Supplementary Figure S4 shows the two confronted dendrograms.

At the genetic level the sample at 1358d (S08) appears to be disruptive between the previous quasispecies and those that follow. The quasispecies in the last four samples are more closely related than expected from the master's UPGMA tree. At the functional level the UPGMA tree of the quasispecies phenotypes is similar in structure to the tree of the master's phenotypes, with larger distances between the last four samples, despite sharing a prevalent master phenotype.

Quasispecies fitness fractions and diversity

The QFF of the quasispecies at the genetic and phenotypic level are confronted in Figure 2, with data in Supplementary Table S2:

At the genetic level, the diversity is very high since the first sample, with most samples showing a fraction of reads for rare and very rare haplotypes above 50%, except for the sample at 1358d (S08). The fraction of reads for very rare haplotypes in the quasispecies increases at each sample, consistent with a mutagenic treatment, except for the sample at 1358d. Finally, the emerging fraction is taking substantial values in the last four samples.

At the functional level we see a different scenario, with master phenotypes above 50% in all samples, except for two. Of note the significant fraction of emerging phenotypes in the last four samples, showing the presence of functional phenotypes different to the master phenotype, which is shared by these samples.

As an alternative view to Figure 2, Supplementary Figure S5 shows the evolution in the master and emerging fractions, side by side, and Supplementary Figure S6 shows the number of rare haplotypes/phenotypes, side by side. The quasispecies diversity of each sample is evaluated in terms of the Hill numbers profile[2] and represented in Supplementary Figure S7 as a

complementary view of quasispecies structure and composition. The last four samples show the highest diversity levels at all q values.

Quasispecies population distance & quasispecies distribution similarity between sequential samples

The changes produced in a quasispecies after short periods of time may be followed by computing the similarity in the haplotype distribution between pairs of sequential samples. Three indices are computed: *Cm*, index of common molecules, *ov*, index of distribution overlap, and *yc*, index of Yue-Clayton[5]. These values are represented in Supplementary Figure S8.

At the genetic level the similarity between haplotype distributions is maintained high until the sample at 1358d (S08), with completely different quasispecies in the four samples that follow. On the other hand, at the functional level, the same differentiation is observed in the sample at 1358d (S08), but the next samples recover moderate values of similarity. These results are consistent with the changes in master sequences and frequencies shown in Figure 1.

The net genetic and phenotypic distances, D_A, between pairs of sequential samples are depicted in Supplementary Figure S9, showing a similar pattern of evolution between the two. The intra-quasispecies diversity of each sample is shown in Supplementary Figure S10, evidencing a big rise after 1358d.

Fraction of synonymous reads and haplotypes to the master phenotype

As commented above, the higher frequencies of the master phenotypes compared to the master haplotypes suggest the presence of multiple haplotypes synonymous to the master phenotype. In studying the fraction of synonymous reads to the master phenotype of each sample (Figure 3), we observe consistent values above 60%, except for two samples, 2065d (40.96%) and 2096d (35.43%). The last sample shows again a fraction above 60% (66.86%). On the other hand, the fraction of synonymous haplotypes to the master phenotype show values above 40% except for three samples; 1358d (30.90%), 2065d (34.35%) and 2096d (30.72%). The last samples increase to the highest value in the series at 57.34%. The ratio of number of haplotypes to number of phenotypes is given in Supplementary Table S3.

High multiplicity synonymous mutants

The relevancy of these synonymous haplotypes is further analysed by aggregating the reads of these synonymous haplotypes according to the number of substitutions with respect to the master haplotype in each sample (Supplementary Figure S4, Supplementary Table S4 and S5). All samples from 0d to 1162d show similar structure, although with increasing fractions at 1 and 2 substitutions. The sample at 1358d shows a prevalent master, followed by a significant fraction

at 1 substitution, and by tiny fractions with 2 to 7 substitutions. These fractions proliferate showing the 1414d sample with a small fraction of mater reads, and increased fractions with 2 to 6 substitutions. The proliferation of these high multiplicity mutants increases mainly in the 5 to 9 fractions in the next two samples. The atomization observed with the samples at 2065d and 2135d causes that the master haplotype expresses a phenotype different to the master phenotype since m0 is null. The last three samples in the series show significant fractions for 10 to 14 substitutions. They have been aggregated together in the m9.up fraction given the limited number of distinguishable colors.

Frequencies of top 10 haplotypes and phenotypes

Beyond the fraction of molecules expressing the master phenotype in each sample, which contribute to a higher frequency in the master phenotype with respect to the master haplotype, some samples show emerging phenotypes with enough functionality, given its frequency. These phenotypes are variants of the main phenotype where some amino acid has been changed by another resulting in a protein of similar chemical structure and functionality. This is particularly notorious in the last four samples (Figure 5).

Amino acid mutations

The probability to observe an amino acid mutation in a protein depends of two factors. The first, is the minimum number of nucleotide substitutions required to produce the mutation. The second is the chemical similitude between the two amino acids, in terms of molecular volume, polarity and chemical composition (Supplementary Figure S11). A high chemical similitude between the wild type amino acid and the new one, will likely not cause significant alterations in the 3D-chemical structure of the protein and in its functionality. In some cases, conservative substitutions may even enhance the function of a protein. For example, a conservative substitution in a binding site may improve the affinity of a protein for its ligand.

The mutations observed in master or emerging phenotypes (frequency above 1%) in the last four samples, with respect to the d0 master phenotype are shown in Supplementary Table S2. The most prevalent changes correspond to T → N and T → A, present in 27 and 23 different phenotypes, respectively. Observed aminoacidic changes require a single nucleotide substitution to occur (distance of Fitch=1). Most of the mutations correspond to a Grantham distance below Q1, five being below D1. These results suggest the conservative nature of these mutations.

Pattern of substitutions

The substitutions with respect to the consensus sequence were analysed in terms of its type and frequency, looking for the pattern of substitutions expected from the treatment with RBV. Most of the substitutions were transitions, as expected, but the substitutions C→T / C and G→ A / G

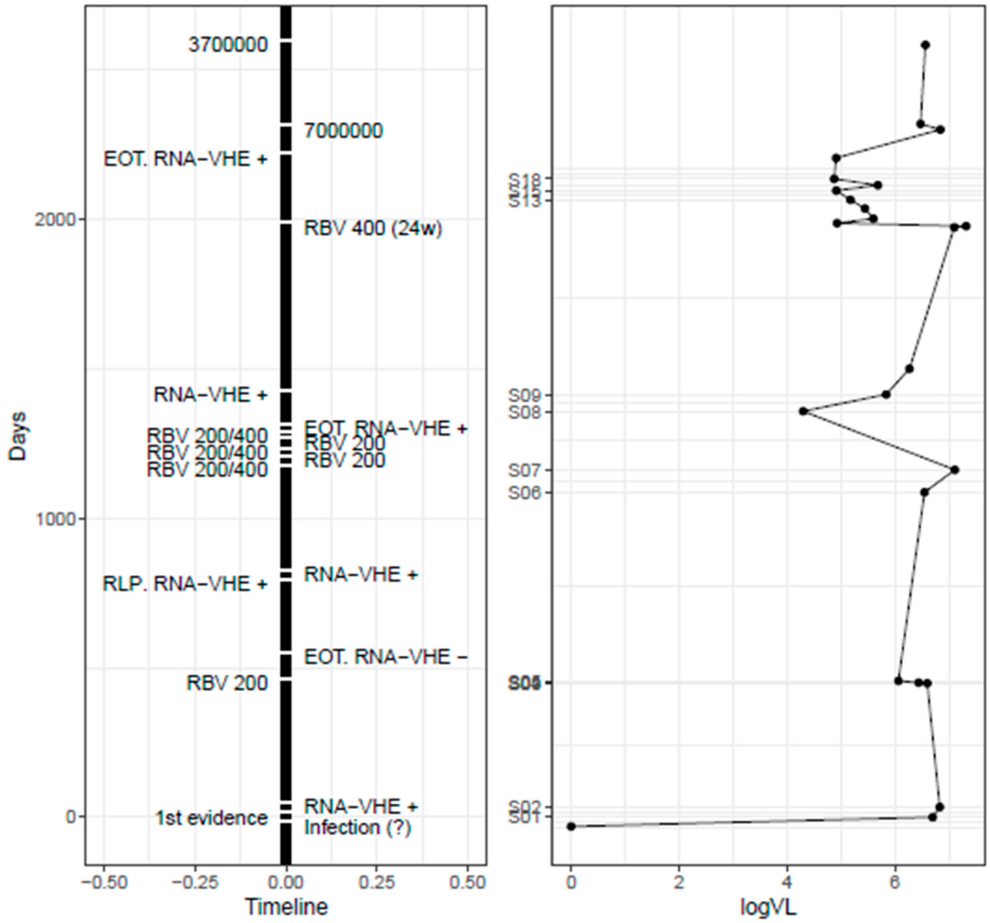
are only clearly prevalent in the last four samples (Supplementary Table S6). In samples where a master haplotype dominates the quasispecies, this analysis would be done computing the substitutions with respect to this master in the hope that most substitutions were produced from the master sequence. In the present scenario, the low frequency of the masters in most samples suggests that a better alternative would be the use of the consensus as reference. Nevertheless, the consensus is an artificial construction, and we may not exclude that its use could somehow blur the results.

REFERENCES

1. Gregori, J.; Soria, M.E.; Gallego, I.; Guerrero-Murillo, M.; Esteban, J.I.; Quer, J.; Perales, C.; Domingo, E. Rare Haplotype Load as Marker for Lethal Mutagenesis. *PLoS One* **2018**, *13*, doi:10.1371/journal.pone.0204877.
2. Gregori, J.; Colomer-Castell, S.; Campos, C.; Ibañez-Lligoña, M.; Garcia-Cehic, D.; Rando-Segura, A.; Adombie, C.M.; Pintó, R.; Guix, S.; Bosch, A.; et al. Quasispecies Fitness Partition to Characterize the Molecular Status of a Viral Population. Negative Effect of Early Ribavirin Discontinuation in a Chronically Infected HEV Patient. *Int J Mol Sci* **2022**, *23*, 14654, doi:10.3390/ijms232314654.
3. Nei, M. *Molecular Evolutionary Genetics*; Columbia University Press: New York, 1987; ISBN 9780231886710.
4. Grishin, V.N.; Grishin, N. V Euclidian Space and Grouping of Biological Objects. *Bioinformatics* **2002**, *18*, 1523–1534, doi:10.1093/bioinformatics/18.11.1523.
5. Gregori, J.; Ibañez-Lligoña, M.; Quer, J. Quantifying In-Host Quasispecies Evolution. *Int J Mol Sci* **2023**, *24*, doi:10.3390/IJMS24021301.
6. Fitch, W.M. An Improved Method of Testing for Evolutionary Homology. *J Mol Biol* **1966**, *16*, 9–16, doi:10.1016/s0022-2836(66)80258-9.
7. Grantham, R. Amino Acid Difference Formula to Help Explain Protein Evolution. *Science* **1974**, *185*, 862–864, doi:10.1126/science.185.4154.862.

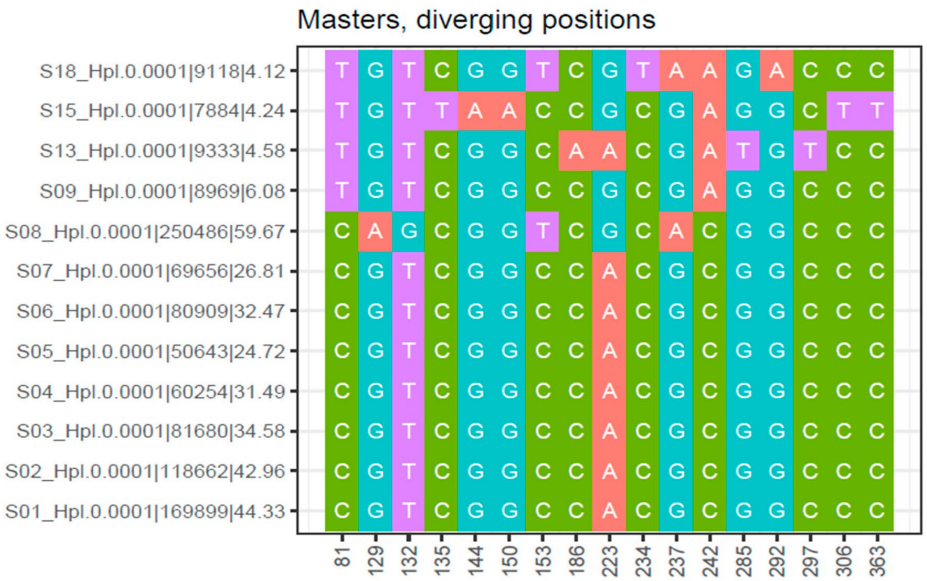
SUPPLEMENTARY FIGURES

Supplementary Figure S1: Time line of interventions, samplings and viral loads, in days since first evidence. Only viral loads corresponding to samples are shown. The first EOT shows negativization of HEV RNA, but is followed by a relapse (RLP) after few weeks.

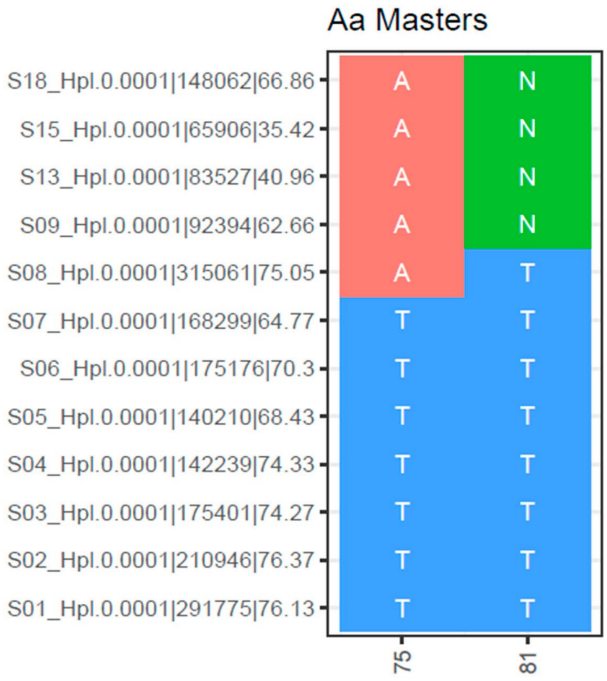


Supplementary Figure S2: Polymorphic sites in the comparison of A) master haplotypes and B) phenotypes. Positions within the amplicon are given in abscissa (nt 6347-6709, aa 409-529).

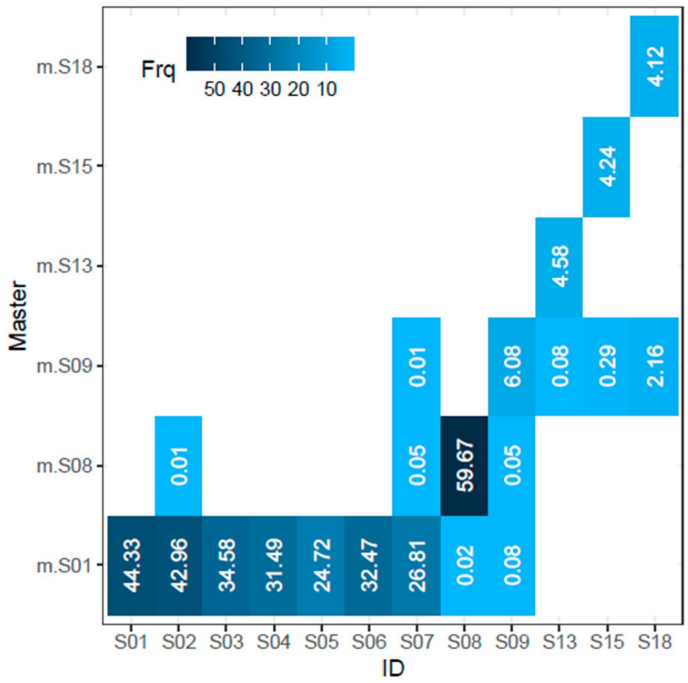
A



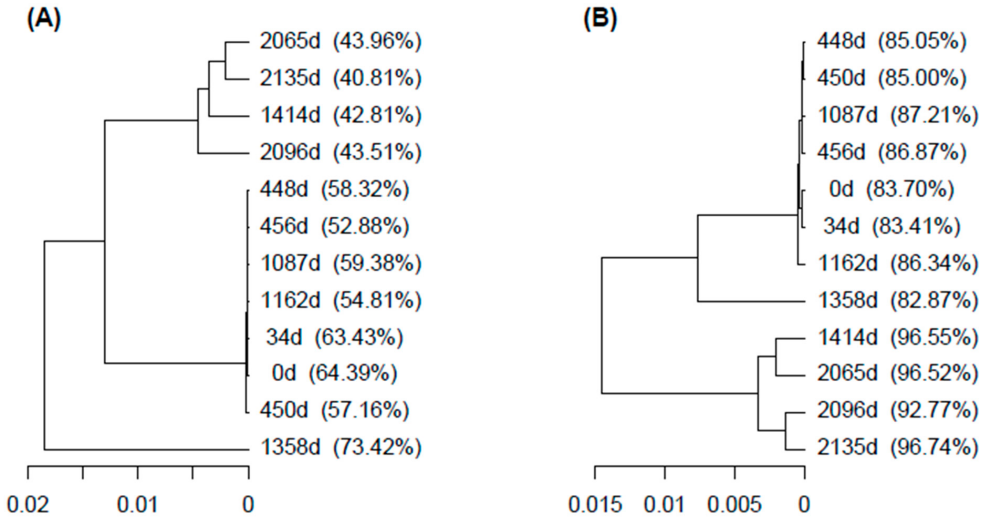
B



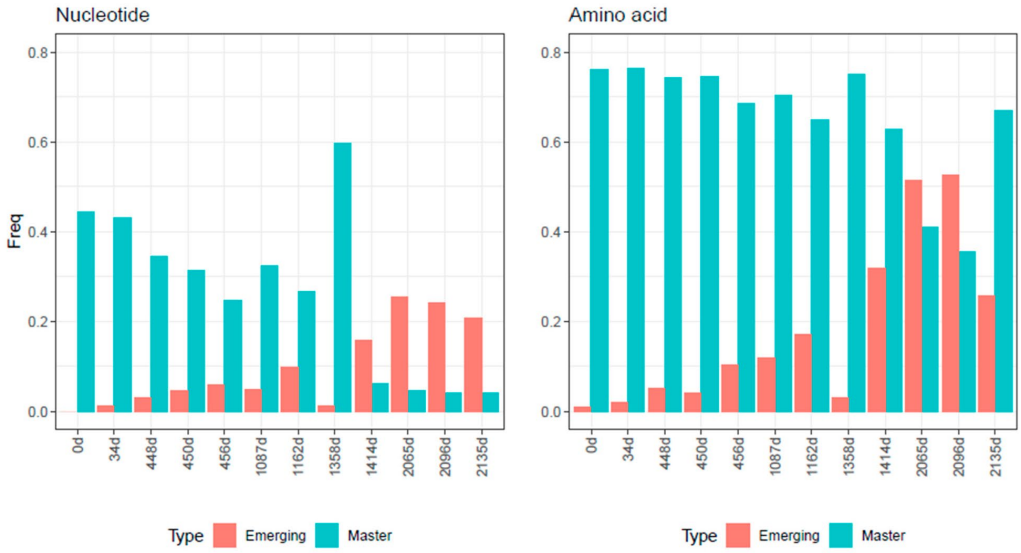
Supplementary Figure S3: Track in frequencies of the master haplotypes in samples evolution.



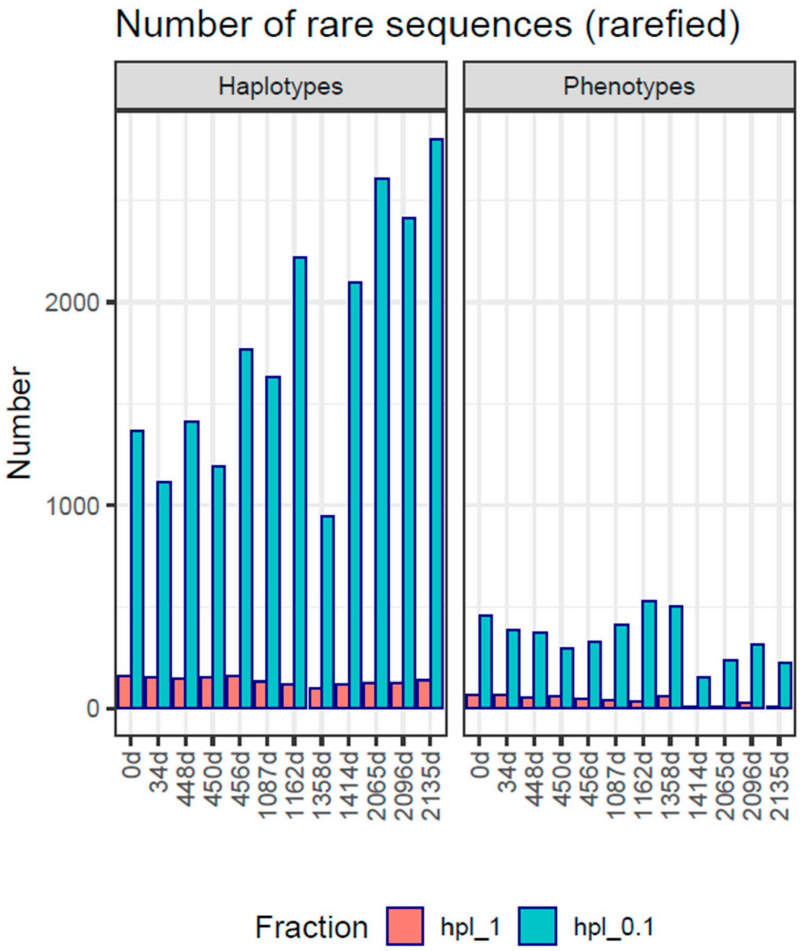
Supplementary Figure S4: (A) UPMA tree with the quasispecies represented by the top 50 haplotypes of all samples. (B) Corresponding tree with the quasispecies represented by the top 20 phenotypes of all samples. The percentages besides the sample labels give the fraction of reads in the quasispecies corresponding to the top 50 haplotypes, or top 20 phenotypes, which were multiple aligned to compute the quasispecies distances.



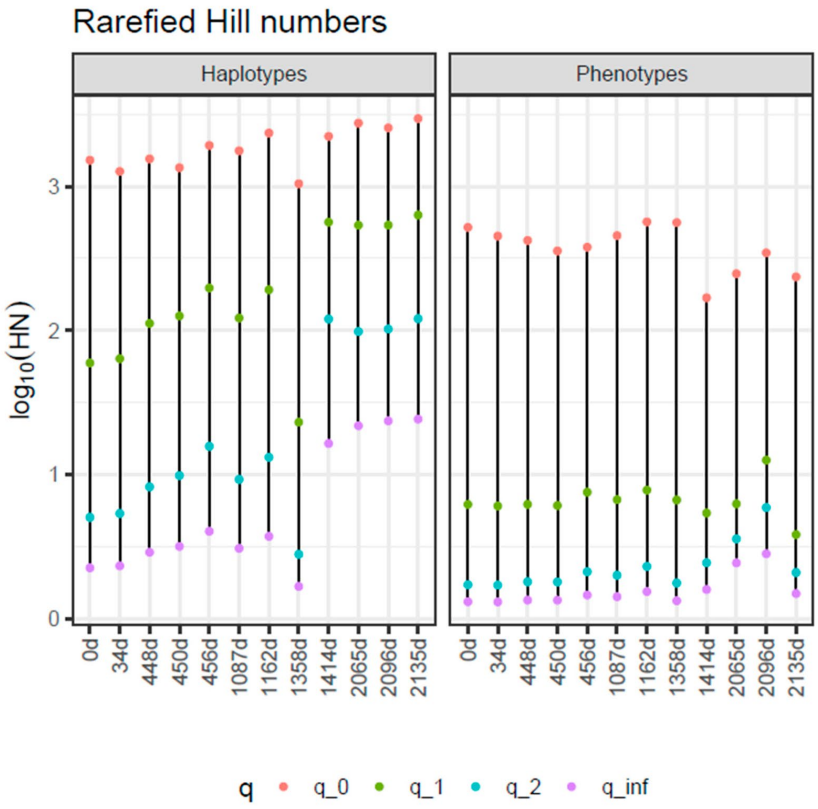
Supplementary Figure S5: Changes in master and emerging fractions. Genetic and phenotypic levels.



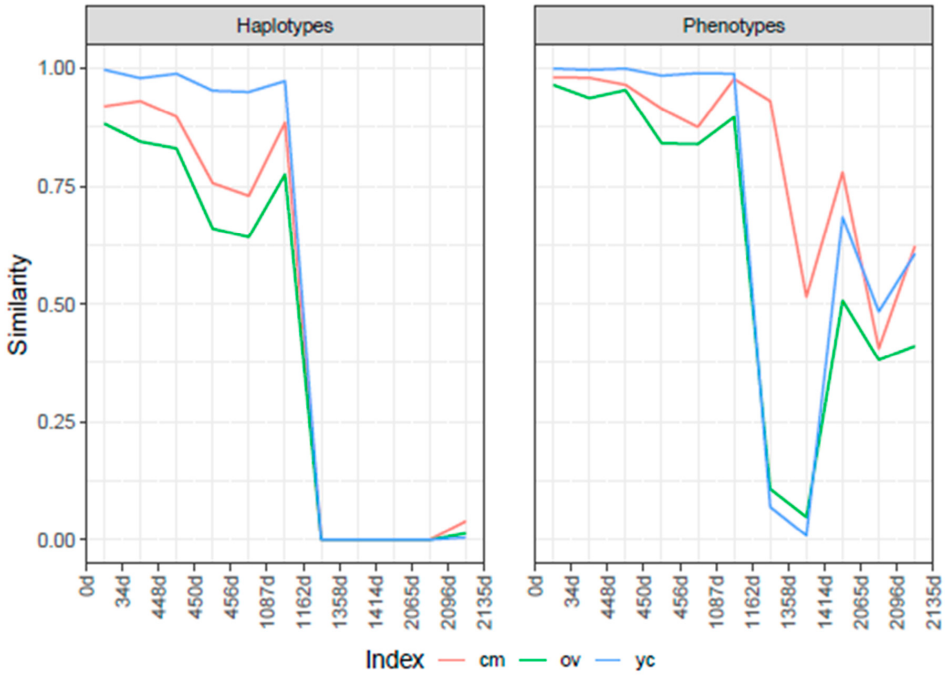
Supplementary Figure S6: Evolution in the number of rare haplotypes. Genetic and phenotypic levels.



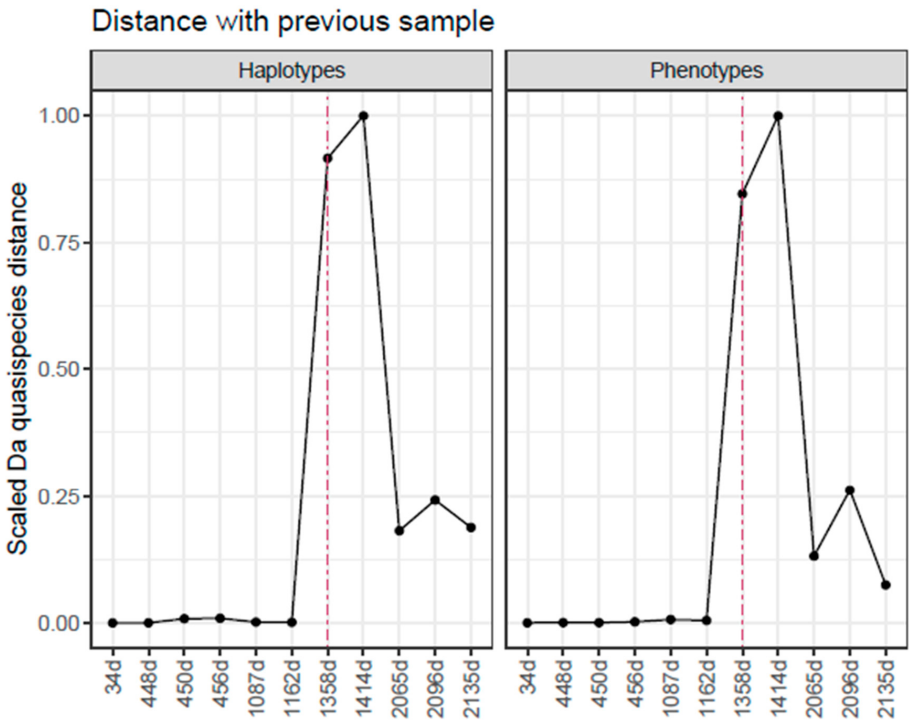
Supplementary Figure S7: Quasispecies diversity as Hill numbers profiles. Only values at q = 0, 1, 2, and Infinity are represented.



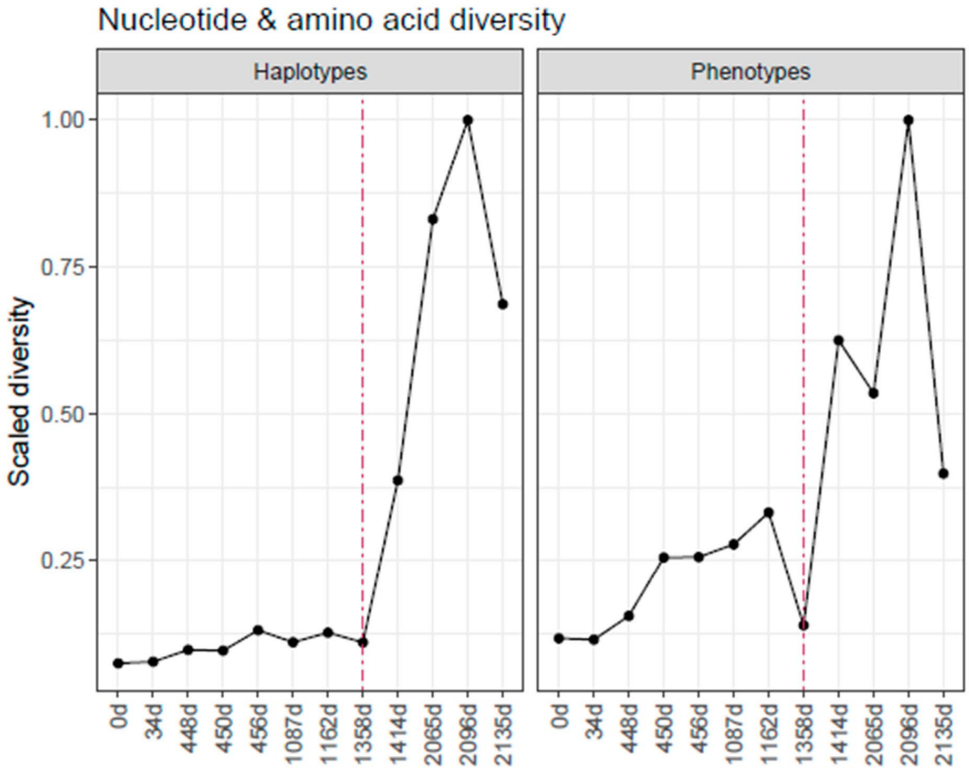
Supplementary Figure S8: Similarities in haplotype or phenotype quasispecies distribution between sequential samples. The 50 top haplotypes, and the 20 top phenotypes are considered. Cm: index of common molecules, ov: index of overlap, yc: index of Yue-Clayton. Similarity values are represented in the mid point between samples.



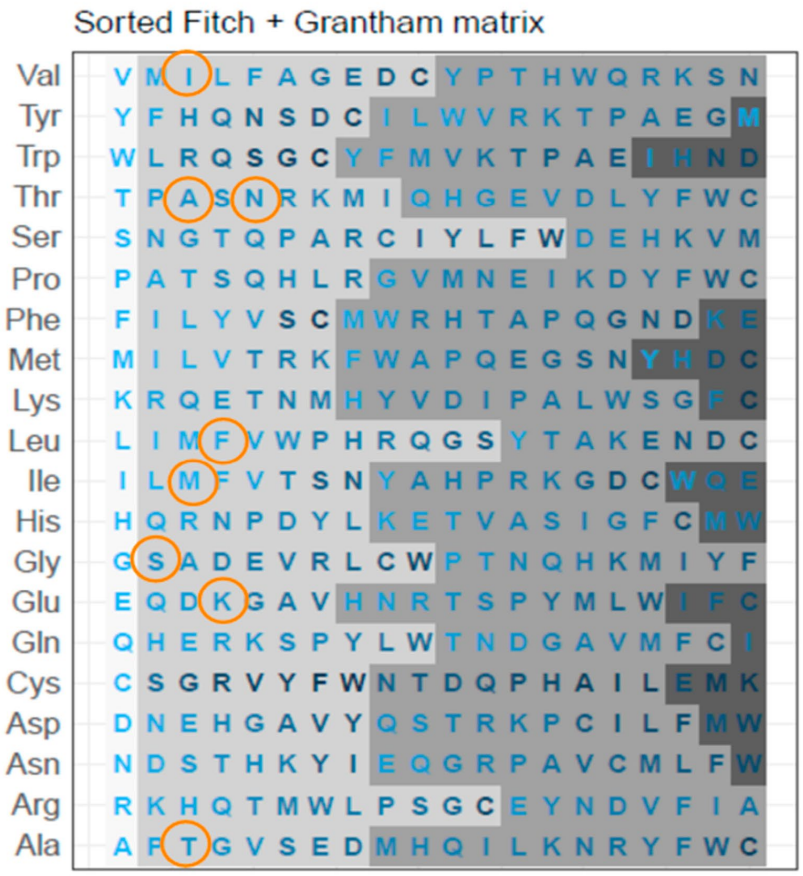
Supplementary Figure S9: Net genetic and phenotypic distance (D_A), between pairs of sequential samples. The 50 top haplotypes, and the 20 top phenotypes are considered. D_A values have been scaled to the maximum value in the sequence to render comparable genetic and phenotypic distances.



Supplementary Figure S10: Nucleotide and amino acid diversity of each sample. Values are scaled to the maximum value in the series to render comparable genetic and phenotypic values.



Supplementary Figure S11: Table of amino acid mutations sorted first by the Fitch distance (gray background) and then by the Grantham distance (blue shade), in increasing order of values. The mutations in Table 2 with N>=2 have been marked with a circle. The mutations nearest to the left, in each row, represent those expected to occur with higher chance and less impact on protein function. The lighter the blue color of the amino acid letter, the smaller the expected impact on protein functionality.



SUPPLEMENTARY TABLES

Supplementary Table S1: Days: days since first evidence, iWks: Interval in weeks between samples; Label: sample label used in figures and table; RBV: regimen of treatment; logVL base 10 logarithm of viral load; nReads: total number of reads in sample; nHpl: number of haplotypes in quasispecies; RdMaster: reads for the master haplotype in each sample.

ID	Date	Days	iWks	Label	RBV	logVL	nReads	nHpl	RdMaster
S01	13 November 2012	0	0	0d	-	6.70	383,263	1,526	169,899
S02	17 December 2012	34	4.9	34d	-	6.84	276,199	1,271	118,662
S03	4 February 2014	448	59.1	448d	200	6.60	236,181	1,554	81,680
S04	6 February 2014	450	0.3	450d	200	6.44	191,367	1,350	60,254
S05	12 February 2014	456	0.9	456d	200	6.07	204,883	1,927	50,643
S06	5 November 2015	1,087	90.1	1087d	-	6.55	249,191	1,768	80,909
S07	19 January 2016	1,162	10.7	1162d	200/400	7.11	259,841	2,348	69,656
S08	2 August 2016	1,358	28.0	1358d	-	4.30	419,804	1,047	250,486
S09	27 September 2016	1,414	8.0	1414d	-	5.84	147,463	2,227	8,969
S13	10 July 2018	2,065	93.0	2065d	400	5.18	203,926	2,751	9,333
S15	10 August 2018	2,096	4.4	2096d	400	4.91	186,055	2,549	7,884
S18	18 September 2018	2,135	5.6	2135d	400	4.88	221,446	2,956	9,118

Supplementary Table S2: Rarefied quasispecies fitness fractions at the genetic and phenotypic levels

	Haplotypes				*Phenotypes*			
ID	Master	Emerging	RHL_1_0.1	RHL_0.1	Master	Emerging	RHL_1_0.1	RHL_0.1
0d	0.4433	0.0000	0.3652	0.1915	0.7613	0.0113	0.1263	0.1012
34d	0.4297	0.0129	0.3588	0.1987	0.7637	0.0203	0.1093	0.1066
448d	0.3458	0.0300	0.3659	0.2583	0.7427	0.0519	0.1006	0.1048
450d	0.3149	0.0438	0.3846	0.2567	0.7433	0.0407	0.1233	0.0927
456d	0.2471	0.0597	0.4317	0.2614	0.6843	0.1037	0.1353	0.0768
1087d	0.3247	0.0538	0.3636	0.2580	0.7029	0.1191	0.0840	0.0940
1162d	0.2680	0.1023	0.2967	0.3330	0.6477	0.1715	0.0667	0.1141
1358d	0.5967	0.0118	0.1846	0.2069	0.7505	0.0299	0.0990	0.1206
1414d	0.0608	0.1589	0.3429	0.4375	0.6266	0.3179	0.0200	0.0355
2065d	0.0458	0.2536	0.2675	0.4331	0.4096	0.5140	0.0381	0.0383
2096d	0.0424	0.2461	0.2701	0.4414	0.3542	0.5268	0.0747	0.0442
2135d	0.0412	0.2061	0.3101	0.4427	0.6686	0.2565	0.0378	0.0370

Supplementary Table S3: Number of haplotypes, phenotypes and the ratio between them, showing the level of synonymous haplotypes in each quasispecies.

ID	*Haplotypes*	*Phenotypes*	*Ratio*
0d	1526	521	2,93
34d	1271	453	2,81
448d	1554	422	3,68
450d	1350	357	3,78
456d	1927	379	5,08
1087d	1768	456	3,88
1162d	2348	568	4,13
1358d	1047	562	1,86
1414d	2227	169	13,18
2065d	2751	248	11,09
2096d	2549	346	7,37
2135d	2956	236	12,53

Supplementary Table S4: Number of haplotypes at increasing number of substitutions with respect to the master haplotype in each quasispecies. Only haplotypes synonymous to the master phenotype are considered.

nHpl	*Substitutions vs master haplotype in synonymous vs master phenotype*														
	0	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**
0d	1	211	465	8	0	0	0	0	0	0	0	0	0	0	0
34d	1	202	377	9	0	0	0	0	0	0	0	0	0	0	0
448d	1	190	547	25	4	0	0	0	0	0	0	0	0	0	0
450d	1	160	521	22	3	0	0	0	0	0	0	0	0	0	0
456d	1	141	521	89	3	2	0	0	0	0	0	0	0	0	0
1087d	1	157	521	24	3	2	0	0	0	0	0	0	0	0	0
1162d	1	185	521	35	2	0	0	0	0	0	0	0	0	0	0
1358d	1	221	521	15	12	9	7	9	0	0	0	0	0	0	0
1414d	1	86	521	372	186	87	29	6	1	0	0	0	0	0	0
1500d	1	97	521	279	103	8	0	0	0	0	0	0	0	0	0
1973d	1	48	521	128	80	12	6	3	2	0	0	0	0	0	0
2065d	0	1	521	10	29	71	119	233	170	150	77	54	19	6	2
2096d	1	48	521	45	67	70	79	109	120	92	70	26	12	5	2
2114d	1	34	521	190	139	106	66	28	17	4	2	0	0	0	0
2135d	0	1	521	60	99	152	319	404	305	175	83	37	20	13	5

Supplementary Table S5: Mutation load. Reads: number of reads in sample; Seqs: Number of unique sequences, genotype and phenotype level. Subst: Total number of substitutions with respect to the master haplotype. SubsLoad: ratio Subst/Seqs; Muts: total number of mutations observed with respect to the master phenotype; MutLoad: ratio Muts/Seqs.

		Haplotypes			*Phenotypes*		
ID	**Reads**	**Seqs**	**Subst**	**SubsLoad**	**Seqs**	**Muts**	**MutLoad**
0d	383.263	1.526	237.021	0,618	521	206.459	0,539
34d	276.199	1.271	173.309	0,627	453	144.478	0,523
448d	236.181	1.554	188.063	0,796	422	110.619	0,468
450d	191.367	1.350	171.584	0,897	357	121.594	0,635
456d	204.883	1.927	219.934	1,073	379	107.134	0,523
1087d	249.191	1.768	221.522	0,889	456	154.246	0,619
1162d	259.841	2.348	260.635	1,003	568	142.748	0,549
1358d	419.804	1.047	305.766	0,728	562	296.364	0,706
1414d	147.463	2.227	394.977	2,678	169	80.240	0,544
2065d	203.926	2.751	1.283.855	6,296	248	163.983	0,804
2096d	186.055	2.549	1.512.102	8,127	346	184.330	0,991
2135d	221.446	2.956	1.431.794	6,466	236	90.838	0,410

Supplementary Table S6: Number of substitutions by type, with respect to consensus sequence in each quasispecies.

	Substitution freq.				*Wild type freq.*			
	C->T	**G->A**	**T->C**	**A->G**	**C**	**G**	**T**	**A**
0d	471	176	1078	292	103	83	98	79
34d	418	181	752	294	103	83	98	79
448d	593	219	1148	449	103	83	98	79
450d	646	214	905	405	103	83	98	79
456d	1135	647	1349	421	103	83	98	79
1087d	886	467	982	474	103	83	98	79
1162d	1108	785	1151	670	103	83	98	79
1358d	353	437	344	331	102	83	98	80
1414d	3115	2877	679	240	102	84	98	79
2065d	4912	5781	1104	537	101	84	99	79
2096d	6607	6194	441	360	101	84	99	79
2135d	6102	5541	587	382	101	84	99	79

5. COMPLEMENTARY RESULTS TO THE COMPENDIUM

5.1 Study 3

Whole-genome Analysis of HEV Under Sequential Ribavirin Pressure Reveals Early Minor Variants Predicting Resistance.

5.1.1 Study 3 Summary

5.1.1.1 Introduction

The longitudinal study of the two HEV chronically infected patients (Study 1 and 2) revealed that failure to RBV treatment generated a highly diverse quasispecies. This quasispecies had a flat-structure with an unstructured quasispecies composed by a master haplotype represented by a very low number of sequences (low frequency) surrounded by a large cloud of mutants at the nucleotide level (genotype). Most of the substitutions were synonymous having a structured quasispecies at the amino acid level (phenotype). Three possible solutions can explain the treatment failure observed during the three different RBV regimens, first of all, is that resistance to RBV is caused by the complex nature of the quasispecies showing no resistance mutations at appreciable frequencies; secondly, that some RBV-resistance substitutions are selected during antiviral treatment; and thirdly, a combination of both in which the extremely high variability includes classical resistance mutations combined with compensatory mutations able to cause at the same time resistance to RBV-treatment together with a fitness increase.

To address this challenge, we extended beyond previous analysis in which only a short genomic region (ORF2) was sequenced, limiting our ability to assess the selection and fixation of resistance mutation across the entire viral genome. By applying untargeted shotgun metagenomics, we successfully sequenced the complete HEV genome, enabling a comprehensive examination of mutation dynamics, and the distribution of the resistance-associated changes and other non-previously reported substitutions throughout the full HEV genome.

5.1.1.2 Objectives

1. To apply shotgun metagenomic sequencing to recover the full-length HEV genome from a chronically infected patient with RBV resistance.
2. To characterize genome-wide variation across sequential RBV treatment courses in order to identify, track, and evaluate the emergence and fixation of RBV resistance-associated mutations.
3. To perform frequency-based deep variant profiling to detect RBV resistance mutations at early stages, and to assess their longitudinal dynamics within the viral quasispecies.

5.1.1.3 Study design

- This longitudinal clinical study focused on a 27 years old patient who developed chronic HEV following two kidney transplantations. The patient underwent three sequential RBV courses:
  - 600 mg/day for three months. Almost achieved HEV RNA negativization.
  - 800 mg/day
  - 1000 mg/day
- Seventeen plasma samples were processed. The viral RNA was extracted.
- Shotgun metagenomics on the Illumina NextSeq2000 platform was performed after Illumina Truseq Total RNA library preparation.
- Samples were analysed in three different run experiments. Reads from the same samples were collapsed to enhance the sensitivity.
- Raw reads were filtered to remove human sequences and technical noise. Variants were identified using LoFreq. Both *de novo* and reference-based assemblies were used to characterize all variants in the viral populations.

5.1.1.4 Results according to objectives

5.1.1.4.1 Objective 1: To apply shotgun metagenomic sequencing to recover the full-length HEV genome from a chronically infected patient with RBV resistance.

The study successfully recovered reads from 16 out of the 17 samples using shotgun metagenomics. Full genome coverage (>90%) was reached in 41% of the samples (7/17), with a mean depth higher than 58x. Consensus sequences were generated via *de novo* assembly for samples with high viral loads (S1, S2 and S6) while highly variable samples (S7, S16 and S17) required reference-based assembly. As a reference-based assembly, we used the S1 full-length HEV genome. The performance for genome recovery was strongly dependent on viral load (**Table 1**).

Table 1. VL in Log (VL) and three different sequencing parameters (CPM, Mean depth and % Covered) are shown for all samples. CPM accounts for counts per millions; % Covered constitutes the percentage of HEV genome covered by sequencing.

Sample	Log (VL)	CPM	Mean depth	% Covered
S1	5.85	68107.73	226.91	100%
S2	5.91	45098.16	798.53	99.98%
S3	5	3172.02	28.08	99.20%
S4	1.54	0	0	0%
S5	3	72.36	0.7	9.13%
S6	4.72	36010.5	135.55	99.97%
S7	6.43	176684.53	1924.38	100%
S8	4.62	1266.5	4.46	71.95%
S9	3.18	426.82	4.14	31.56%
S10	2.04	5.3	0.04	2.50%
S11	2	334.22	0.06	3.87%
S12	2.48	199.01	0.16	8.18%
S13	3.04	740.85	3.17	27.28%
S14	3.4	1348.95	6.05	40.22%
S15	3.68	519.35	2.37	37.93%
S16	4.45	20871.38	58.7	96.73%
S17	6.34	763104.09	4720.61	100%

5.1.1.4.2 Objective 2. To characterize genome-wide variation across sequential RBV treatment courses in order to identify, track, and evaluate the emergence and fixation of RBV resistance-associated mutations.

The analysis of genome-wide variations tracked the accumulation of mutations under sequential RBV pressure. The number of identified nucleotide variants followed significant upward trend, increasing from 33 in the baseline sample S1 to 603 in the final sample S17 (**Figure 16**, corresponding to figure 2 in Article 3). In addition, transition to transversion (Ts/Tv) ratio evolved from 7.25 to 36.7, reflecting the mutagenic effect of successive RBV treatments.

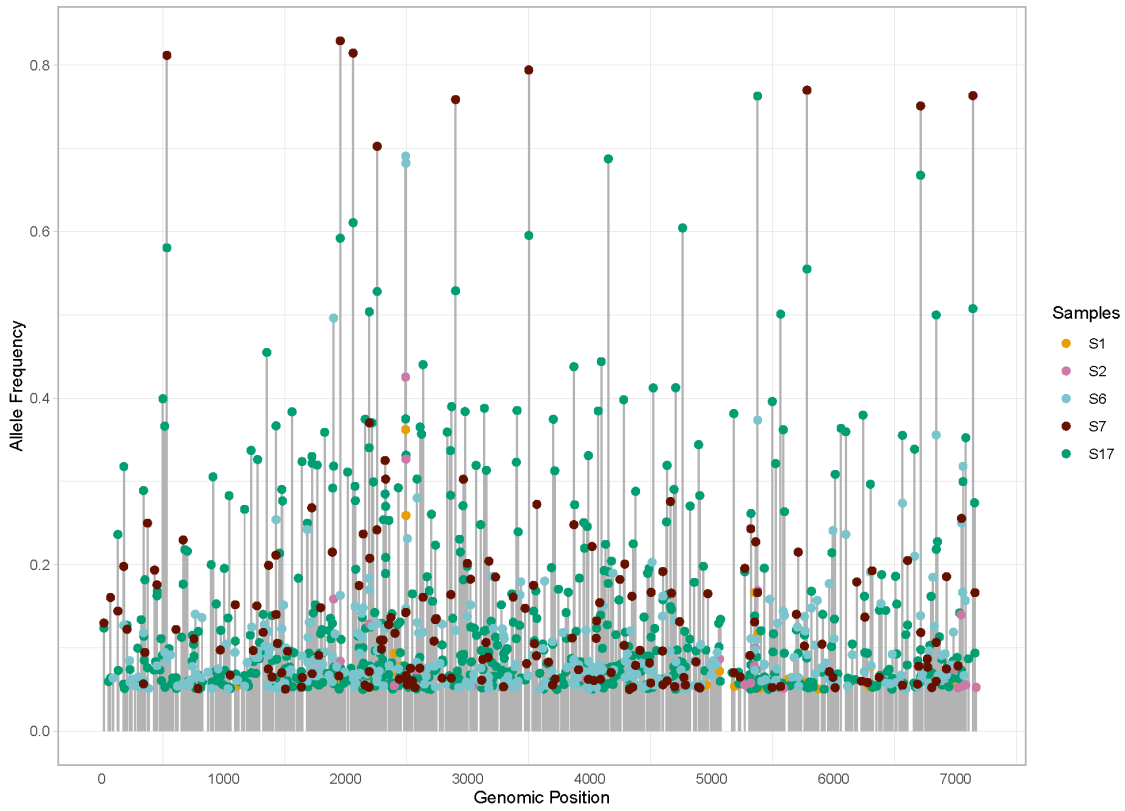


Figure 16. Lollipop Chart showing the allele frequency (Y axis) from all the variants identified in S1, S2, S6, S7 and S17. Variants are distributed following the genome position (X axis). All variants were filtered from 0.05 AF and 50 reads.

Regarding the mutations fixed at the consensus level (AF>0.5), 11 amino acid mutations were identified. Notably, the known RBV resistance mutations D1384N and Y1587F, located in the RdRp domain, that became fixed in the final sampling point S17 (**Table 2**).

Table 2. Mutations occurring at the consensus level from S2, S6, S7 and S17 in pORF1, pORF2 and pORF3. Its presence between consensus threshold and AF < 0.05 is showed in the right column.

Protein	Domain	Mutation	Consensus	Presence at AF > 0.05
pORF1	**PCP-HVR InterDomain**	G643E	S6	S2, S6, S17
		R661H	S7, S17	S2, S6, S7, S17
	HVR-PRP	W738R	S17	S1, S2, S6, S7, S17
		E755G	S7, S17	S6, S7, S17
	X	H830Y	S6	S1, S2, S6, S7, S17
		P831S	S6	S1, S2, S6, S7 (P831W), S17
	RdRp	D1384N	S17	S17 (D1384T)
		Y1587F	S17	S17
pORF2	**P**	Y532D	S7, S17	S6, S7, S17
		I644V	S6	S6, S7, S17
pORF3		P90L	S17	S1, S2, S6, S7, S17

5.1.1.4.3 Objective 3: To perform frequency-based deep variant profiling to detect RBV resistance mutations at early stages, and to assess their longitudinal dynamics within the viral quasispecies.

Deep variant profiling was approached by addressing mutations with an allele frequency > 0.05 and a depth with at least 50 reads.

Differently from the fixed mutations at the consensus level, a total of 202 mutated positions were found, 140 occurring in ORF1, 44 in ORF2 and 18 in ORF3. HEV RdRp mutations previously reported in literature were also

detected in S17, including K1383N and G1634R. In the case of G1634R, it already appeared in S7, when the second RBV 800mg treatment was started. To assess whether additional mutations were present at low haplotype frequencies, we examined variants with AF < 0.05 and found that D1384N was already detectable in S7, while K1398R only appeared in S17 after three RBV treatments.

Taken together, most amino acid substitutions were detectable at sub-consensus frequencies, revealing an expanding mutant cloud that gradually incorporated higher-fitness substitutions over time.

The early detection of resistance long before those mutations became dominant in the viral population would have avoided unnecessary drug exposure, especially in the case of HEV treated with RBV. The study revealed that a high overall mutational load in later samples can mask key substitutions at the consensus level, making deep profiling essential to reveal the complete state of the quasispecies.

5.1.1.5 Conclusions

- HEV exhibits significant genomic plasticity under sequential RBV pressure, with viral diversity and transition to transversion (Ts/Tv) ratio increasing over time.
- Resistance variants to RBV could be detectable as minor variants at sub-consensus level at the moment of giving the second RBV treatment after the first viral rebound, long before they became dominant in the consensus sequence.
- HEV might be able to adapt to ribavirin by exploiting the variability introduced and the expansion of the mutant cloud generated. The larger the mutant pool, the greater the probability of a resistance mutation emerging.
- Relying solely on consensus sequence analysis is insufficient. Deep variant profiling provides significant clinical value for anticipating RBV treatment failure. Precision medicine approaches could enable clinicians to make more informed decisions and avoid ineffective treatments.

5.1.2 Study 3

Whole-genome Analysis of HEV Under Sequential Ribavirin Pressure Reveals Early Minor Variants Predicting Resistance.

Sergi Colomer-Castell, Marta Ibañez-Lligoña, Carolina Campos, Josep Gregori, Damir Garcia-Cehic, Ariadna Rando-Segura, Roser Ferrer, Sofía P. Ruzo, Maria Francesca Cortese, David Tabernero, Judit Vico-Romero, Juan Carlos Ruiz-Cobo, Maria Buti, Mar Riveiro-Barciela and Josep Quer*

Whole-genome Analysis of HEV Under Sequential Ribavirin Pressure Reveals Early Minor Variants Predicting Resistance.

Sergi Colomer-Castell[1,2,3], Marta Ibañez-Lligoña[1,3,4], Carolina Campos[1,2,3], Josep Gregori[1], Damir Garcia-Cehic[1,2,5], Ariadna Rando-Segura[3,6], Roser Ferrer[4,7], Sofía P. Ruzo[1], Maria Francesca Cortese[1,3,6], David Tabernero[1,3,6], Judit Vico-Romero[1], Juan Carlos Ruiz-Cobo[1,3,4], Maria Buti[1,3,4], Mar Riveiro-Barciela[1,3], Josep Quer[1,2,3,4,*].

[1]Liver Diseases-Viral Hepatitis, Liver Unit, Vall d'Hebron Institut de Recerca (VHIR), Vall d'Hebron Hospital Universitari, Vall d'Hebron Barcelona Hospital Campus, Barcelona.

[2]Biochemistry and Molecular Biology Department, Universitat Autònoma de Barcelona (UAB), Bellaterra.

[3]Centro de Investigación Biomédica en Red de Enfermedades Hepáticas y Digestivas (CIBERehd), Madrid.

[4]Medicine Department, Universitat Autònoma de Barcelona (UAB), Bellaterra.

[5]Department of Radiation Oncology, University of Cincinnati, College of Medicine, Cincinnati, Ohio, USA

[6]Microbiology Department, Vall d'Hebron Institut de Recerca (VHIR), Vall d'Hebron Hospital Universitari, Vall d'Hebron Barcelona Hospital Campus, Universitat Autònoma de Barcelona, Barcelona.

[7]Biochemistry Department, Vall d'Hebron Institut de Recerca (VHIR), Vall d'Hebron Hospital Universitari, Vall d'Hebron Barcelona Hospital Campus, Barcelona.

***Corresponding author:**

Josep Quer josep.quer@vhir.org

Abstract

Ribavirin (RBV) failures in chronic hepatitis E virus (HEV) infection may arise from genomic adaptation, yet the contribution of minor variants remains insufficiently explored. Using a shotgun metagenomics based on next-generation sequencing to obtain the full HEV genome, we analysed longitudinal HEV populations from sequential clinical samples collected from a patient who underwent three different courses of RBV treatment. Viral diversity increased over time, with most amino-acid substitutions detected at sub-consensus frequencies. Several mutations linked to RBV resistance emerged under treatment pressure. Specifically, D1384N and Y1587F became fixed in the final sampling, while additional substitutions at position 1384 (including D1384T) revealed mutational hotspot. Importantly, D1384N and G1634R were already detectable at low allele frequencies after the first treatment cessation and subsequent rebound, showing that the study of minor variant populations can be early predictors for the development of RBV resistance. These findings underscore the genomic plasticity of HEV under antiviral pressure and highlight the value of deep variant profiling for anticipating RBV treatment failure.

Introduction

Paslahepevirus balayani, commonly known as the hepatitis E virus (HEV), is the causative agent of Hepatitis E, an inflammatory liver disease and the leading cause of acute hepatitis worldwide, which affects approximately 20 million people annually[1]. It is single stranded RNA virus belonging to the Hepeviridae family, genus Hepeviridae, and distributed in 8 different genotypes[2]. Among these, only HEV1 and HEV2 are exclusively affecting humans, while the other ones have zoonotic origin. The distribution of the different genotypes is geographically dependent, being HEV1 present mostly in Southern Asia and African countries following an hyperendemic pattern. HEV2 is also related to an endemic situation in Western African countries and Mexico. These two genotypes are predominantly associated with developing countries, where Hepatitis E represents a major public health concern. Transmission typically occurs via the fecal-oral route, primarily through contaminated water or food, as well as through person-to-person contact. On the other hand, HEV3 is related to sporadic worldwide-distributed infections in developed countries, including Europe, North America, Australia, and some South American countries. HEV4 follows the same pattern but is predominant in China and Japan[3]. HEV3 and HEV4 cause a foodborne hepatitis related to undercooked pig and boar food mostly[4,5]. All this four HEV genotypes can also be transmitted by parenteral route and have been associated to blood and blood derivatives transfusions.

HEV infection usually presents as an acute and transient self-limited hepatitis, with an incubation period of 5 to 6 weeks in average, that rarely leads to acute liver failure[6]. HEV infection, especially HEV1 genotype, could be dangerous for pregnant women, with a maternal mortality rate of 20-30% and risk of fetal loss[7]. Extrahepatic manifestations have been well documented, including neurologic tropism related to Guillain-Barré syndrome, Parsonage-Turner syndrome, and other neuropathies[8]. Moreover, renal manifestations have been described, including disorders such as pre-renal failure, glomerular disorders and tubular and interstitial injuries[9]. In immunocompromised hosts, including organ transplant recipients, HIV or chemotherapy patients, HEV may progress to a chronic stage due to lack of immune clearance of the virus. This phenomenon occurs mainly with HEV3 but has also been reported for HEV4[10].

First-line management typically involves a reduction of immunosuppressive therapy, as recommended by the European Association for the Study of the Liver (EASL)[11]. If the infection persists, ribavirin (RBV) monotherapy is administered for three months, either as a weight-adjusted dose or based on glomerular filtration rate. In cases of treatment failure or relapse, re-treatment with RBV for six months is indicated. In a previous study we demonstrated that monotherapy treatment with RBV can lead to a diversification of the viral population, making the virus more resistant to future treatments. This study led to a change in the RBV-based treatment strategy, requiring therapy continuation for several months beyond plasma viral load clearance, and the implementation of highly sensitive methods to detect viral RNA in both plasma and stool samples. These measures aim to prevent the rebound and diversification previously observed[12]. Alternatively, PEG-Interferon-α has shown efficacy in liver transplant recipients but is contraindicated in other solid organ transplant patients due to its immune-stimulating effects. Alternative therapeutic strategies, mostly involving validated drugs, such as combining ribavirin with sofosbuvir, have been explored but with limited success, underscoring the need for new therapeutic approaches[13].

The lack of proof-reading activity of the viral polymerase, together with the high replication rate typical of RNA viruses, makes HEV incorporating mutations at each replication cycle. Because of that, the virus circulates as a complex and dynamic viral population, formed by several variants, and defined as a quasispecies[14,15]. As an example, most of the RBV resistance mutations reported occur inside the RNA dependent RNA polymerase (RdRp) domain of the ORF1 polyprotein altering the virus replication dynamics[16,17]. Hence, the assessment of the complete viral population is important to reveal viral escape mechanisms.

In our previous studies involving HEV amplicon sequencing a short region in the open-reading frame 2 (ORF2) viral genome, we applied a novel method for studying viral variability, the Quasispecies Fitness Fraction (QFF), which consists in considering the viral populations as fractions based on the variant's frequency[18]. We reported an extreme variability at the nucleotide level probably enhanced by the duration of the infection and the mutagenic effect of RBV treatment. Nevertheless, a conservation at the amino acid level was observed, which might suggest that high diversification of a quasispecies generates a resistant population to further treatments[12,18] . However, this study centered in a short fragment impaired the analysis of the selection and fixation of resistance mutations during the RBV treatment, and especially once the virus has developed resistance to increased doses of RBV.

In the present work, shotgun metagenomics (mNGS) covering the whole HEV genome approach has been followed to identify mutations and follow their dynamicity in a 3-year chronic HEV infected patient. Notably, this study shows that quasispecies diversity at the nucleotide level may be also accompanied by specific amino acid substitutions that can be selected and fixed during RBV treatment. Early detection of these resistant mutations, even when present at low frequency, following first treatment failure and viral load recovery, would provide essential information to avoid unnecessary drug exposure and further selection and fixation of resistance-associated mutations.

Results

Clinical metagenomics NGS for HEV genome recovering

Sequential samples were obtained from a patient that underwent three different RBV regimes (**Figure 1**). Files for the analysis were obtained from by collapsing three different NGS experiments into a single file (**Table 1**). We were able to detect HEV reads for most of the longitudinal samples (16 out of 17), with more than 75% percent of the samples containing >10 CPM (Counts Per Million). The percentage of viral genome covered was strongly depending on CPM and depth, achieving more than 90% in 7/17 (41%) of the analysed samples, with mean depth higher than 58x. Within these quality parameters, we were able to generate HEV consensus sequence from either *de novo* or reference-based assembly. From samples S1, S2, and S6, showing coverages higher than 90% and mean depth over 50x, consensus *de novo* sequences were obtained. On the other hand, although the S7, S16 and S17 presented the same sequencing parameters, their reads were highly variable, making the *de novo* assembly unfeasible. In consequence, for these samples, a consensus sequence from the first tested samples (S1) was used as reference. The genome

recovery performance here presented is consistent with previous data from clinical metagenomics, where we were able to recover the complete genome from samples with viral load (VL) >4 Log (***In Press***). Finally, although S3 and S8 showed the minimum viral load, both had a mean depth <50x, and S8 less than 75% of the genome covered. Because of that, these two samples were not included in further analysis.

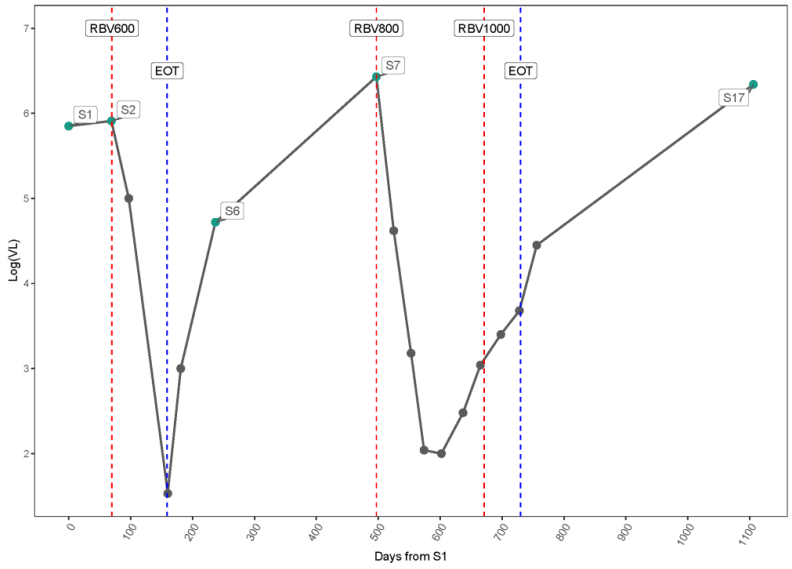


Figure 1. Infection course diagram showing the extraction date of the sample and its corresponding VL in logarithmic scale. Vertical dotted lines indicate the RBV treatment application (red) or End of treatment (Blue). Samples (S1, S2, S6, S7 and S17) with complete genome coverage are highlighted.

Table 1. VL in Log (VL) and three different sequencing parameters (CPM, Mean depth and % Covered) are shown for all samples. CPM accounts for counts per millions; % Covered constitutes the percentage of HEV genome covered by sequencing.

Sample	Log (VL)	CPM	Mean depth	% Covered
S1	5.85	68107.73	226.91	100%
S2	5.91	45098.16	798.53	99.98%
S3	5	3172.02	28.08	99.20%
S4	1.54	0	0	0%
S5	3	72.36	0.7	9.13%
S6	4.72	36010.5	135.55	99.97%
S7	6.43	176684.53	1924.38	100%
S8	4.62	1266.5	4.46	71.95%
S9	3.18	426.82	4.14	31.56%
S10	2.04	5.3	0.04	2.50%
S11	2	334.22	0.06	3.87%
S12	2.48	199.01	0.16	8.18%
S13	3.04	740.85	3.17	27.28%
S14	3.4	1348.95	6.05	40.22%
S15	3.68	519.35	2.37	37.93%
S16	4.45	20871.38	58.7	96.73%
S17	6.34	763104.09	4720.61	100%

Addressing HEV genomic diversity

To address nucleotide variations that appeared at each genome position the generated S1 consensus sequence was used as standard reference. Reference-based mapping was applied to obtain the variants in the five time points include in the analysis after covering the whole HEV genome: S1, S2, S6, S7 and S17. To achieve minimum quality criteria according to the ones previously described (depth and coverage), variants were filtered to be present in at least 50 reads with an allele frequency (AF) higher than 5%.

At the nucleotide level all variants obtained were represented in a Lollipop Chart (**Figure 2**). Most of the variants corresponded to transversions, with transition to transversion (Ts/Tv) rates that evolved from 7.25 in S1 to 36.7 in S17. The number of variants identified followed an increase tendency from S1 to S17 (33, 30, 321, 173 and 603 respectively), related to time of the infection and treatment application. However, the higher mean AF variants were located at position S7 instead of S17, with a value of 0.16 in front of 0.14. Despite this, both mean and IQR AF calculations showed the same upward tendency (**Suppl. Figure 1**).

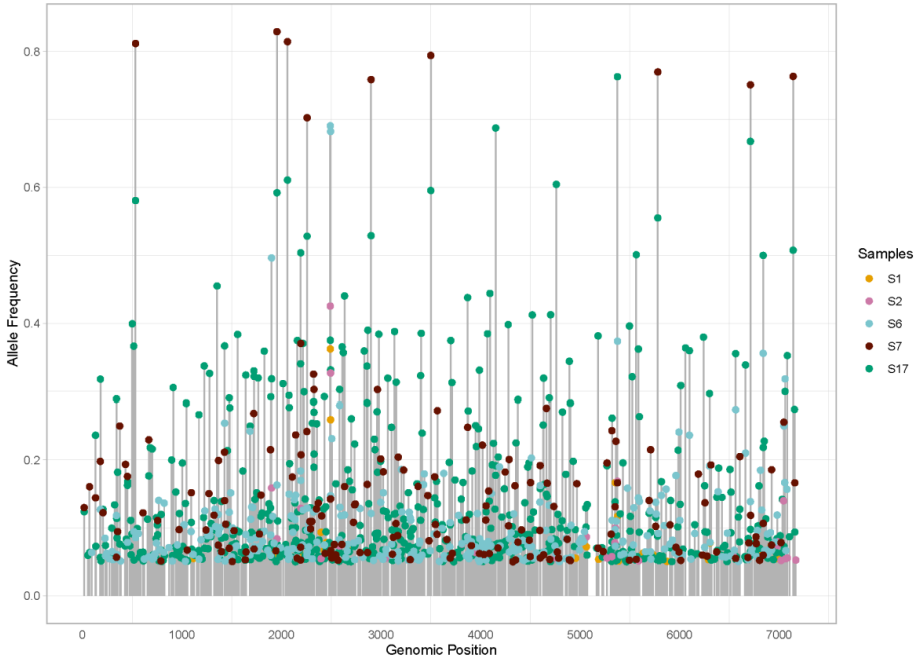


Figure 2. Lollipop Chart showing the allele frequency (Y axis) from all the variants identified in S1, S2, S6, S7 and S17. Variants are distributed following the genome position (X axis). All variants were filtered from 0.05 AF and 50 reads.

Variants evolution over time

To characterize variant emergence and fixation across the time, all sample reads (including samples with uncompleted genome but with a deep coverage in the substitution positions) were aligned to the S1 consensus sequence. Consistent with the previous criteria, only variants supported with ≥50 reads and displaying an AF above 5% were retained. Across all the variants studied, we identified 71 nucleotide positions, that were present in at least three different points, maintaining the same substitutions (**Figure 3**). It is noteworthy that 69 of the variants that appeared along the infection were present at the last point S17, including 7 of the 8 already present in the first sample S1. Forty-eight out of the 71 substitutions (66%) were identified in S6 when the first viral load rebound occurred after the stop of the first RBV treatment. This burden of mutations is accompanied with a rise of the AF mean, with a peak after S14, coinciding with the increase in the RBV dosage to 1000mg in the third treatment. However, AF slightly decreases in S17, coinciding with the general variation results.

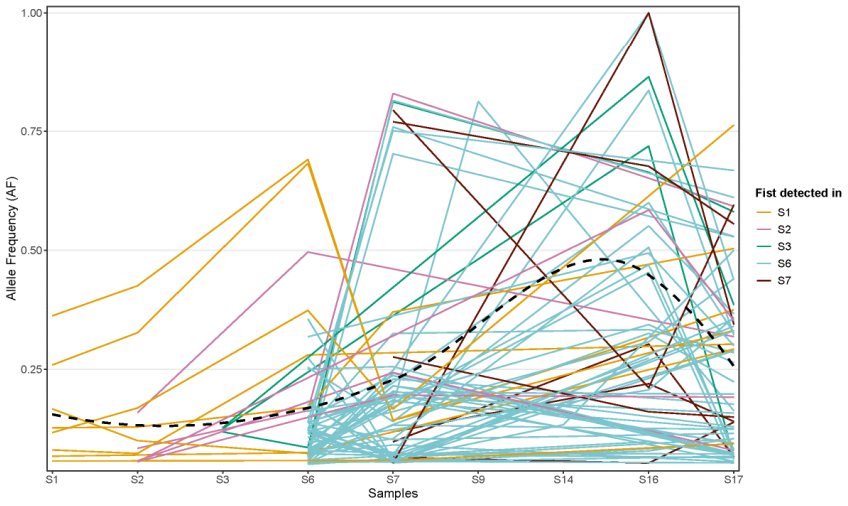


Figure 3. Line point diagram showing the nucleotide mutations that were present in three study points or more. Y axis corresponds to mutation allele frequency while X axis shows the samples where the mutations are present. Colours represent the sample in which the mutation was first detected.

Elevated nucleotide diversity contrasts with low aminoacidic substitution fixation

As previously reported, in the three first samples (S1, S2 and S6) consensus nucleotide sequence was generated through *de novo* assembly, but for S7 and S17, due to the enormous variability at the nucleotide level, it required the use of S1 as reference sequence to generate the assembly. Only variants with an AF > 0.5 were called, ensuring that the substitutions represented fixed changes within the viral population. HEV protein sequences (ORFs) were identified using NCBI's ORF Finder (RRID:SCR_016643; https://www.ncbi.nlm.nih.gov/orffinder)[19]. To account for differences in mutation annotation due to genotype, the Burma strain[20] was included in each alignment for consistent protein position numbering, particularly in ORF1. All alignments were performed using the S1 consensus sequence as reference.

Eleven amino acid mutations were identified (**Table 2**), affecting viral populations present in S6, S7 and S17. Interestingly, no mutations were identified in the S2 sample, corresponding to the RBV starting treatment when virus showed sensitivity to this treatment. Once the first treatment was stopped, the relapsed quasispecies (S6) presented 5 selected substitutions that surprisingly, were not maintained in S7 or S17 after stopping RBV treatment with 1000mg. In contrast, the three mutations fixed in S7 sample were also maintained in S17. S17 represented the viral population with the highest number of amino

acid changes showing 7 out of the 11 substitutions detected after the three RBV treatment regimes.

Mutations in pORF1 were distributed across four different domains. Both G463E and R661H were located in the interdomain hinge between the papain-like cysteine protease (PCP) and the hypervariable region (HVR-PRP). The HVR incorporated W738R and E755G in the later sampling points. S6 contained two consecutive mutations within the X domain (H830Y and P831S). Two key mutations were identified at consensus level in the RNA-dependent RNA-polymerase domain (D1384N and Y1587F) both fixed in the final sample S17.

In pORF2, two mutations were observed: Y532D in S7, which was maintained in S17, and I644V in S6, both located in the P domain of the capsid protein. Only one mutation was identified for pORF3 (P89L in S17). Despite the lower number of mutations, the percentage of mutated positions was nearly equivalent for ORF1 and ORF2, ranging between 0.12% and 0.3% of all the amino acids.

Table 2. Mutations occurring at the consensus level from S2, S6, S7 and S17 in pORF1, pORF2 and pORF3. Its presence between consensus threshold and AF < 0.05 is showed in the right column.

Protein	Domain	Mutation	Consensus	Presence at AF > 0.05
pORF1	**PCP-HVR InterDomain**	G643E	S6	S2, S6, S17
		R661H	S7, S17	S2, S6, S7, S17
	HVR-PRP	W738R	S17	S1, S2, S6, S7, S17
		E755G	S7, S17	S6, S7, S17
	X	H830Y	S6	S1, S2, S6, S7, S17
		P831S	S6	S1, S2, S6, S7 (P831W), S17
	RdRp	D1384N	S17	S17 (D1384T)
		Y1587F	S17	S17
pORF2	**P**	Y532D	S7, S17	S6, S7, S17
		I644V	S6	S6, S7, S17
pORF3		P90L	S17	S1, S2, S6, S7, S17

Amino acid diversity is largely driven by low-frequency variants

Amino acid substitutions present at a level lower than consensus were addressed by modifying bcf tools parameters to filter AF > 0.05 mutations in the generated sequences

for S1, S2, S6, S7 and S17 using S1 consensus sequence as the reference. Is this study, amino acid substitutions with frequencies lower than 0.05 were not included to avoid bias in the analysis due to sequencing or technical errors. ORFs were generated as previously detailed using Burma strain[20] mutation annotation.

A total of 202 mutated positions were found, 140 occurring in ORF1, 44 in ORF2 and 18 in ORF3 (**Suppl. Table S1; S2; S3**). Considering the ORF size, ORF3 was the most mutated in proportion, ranging from 3.54% to 9.73% amino acid point substitutions from S1 to S17 respectively. What is more, the proportion of mutated positions rose over time, in line with treatment progression and in concordance with the nucleotide-level variability described above, although S7 deviated from this overall trend (**Suppl. Figure S2**).

Distribution of the mutations was asymmetrically related to protein domains (**Suppl. Figure S3; Suppl. Figure S4**). Amino acid substitutions were distributed along ORF1 affecting all protein domains, showing a higher concentration in PCP and specifically in the hypervariable region (HVR-PRP). Interestingly, RdRp domain also exhibited a high number of mutations. On the other hand, pORF2 mutations were concentrated in both ends of the proteins, displaying the protein N-terminal region and the P domain the largest number of substitutions.

Interestingly, ORF1 showed 16 mutations that were present in more than 3 samples, which included all the mutations previously described fixed in the consensus **(Table 2**). What is more, W738R, E755G, H830Y and P831S were shown in the viral population since the starting time point S1, although their frequency fluctuated and in some points were not present in the consensus sequence. R661H, appeared at S2 and maintained until S17 (with a frequency reduction in S7). G810C, on the other hand, appeared in S1 but intermittently disappeared from the AF > 0.05 threshold to finally reappear in S17. Finally, the remaining amino acid mutations were selected in S6 and maintained in S17.

pORF2 showed 5 positions with amino acid substitutions in at least 3 of the samples. Three out of five (Y532D, P603S and I644V), appeared in S6 and were present until S17. Both Y532D and I644V were already in the consensus sequence analysis but they were below the AF >0.5 threshold at the other points. On the other hand, mutation P79L appeared in S1 but was not selected and eliminated from the viral population after S7. Lastly, it is interesting to note that P68S, which appeared in S2, dropped in frequency in S6 with AF<0.05 but reappeared in S7 and S17.

pORF3 showed 3 positions with substitutions in more than three samples, highlighting P90L that was present from S1 to S17. Conversely, H83Y was negatively selected,

disappearing in S7 and S17. S55L, on the other hand, appeared in S2 and persisted until S17, although it went below the AF < 0.05 threshold in S6.

To study the impact of the identified amino acid substitutions, the frequency and chemistry of the substitutions was assessed and is shown in **Figure 4**. The most frequent biochemical change in ORF1 and ORF2 was the non-polar to polar substitution, followed by changes that involved only non-polar amino acids. In both ORF1 and ORF2, the most prevalent change was A->T. The V->I substitution also exhibited an elevated frequency, being the second in ORF1 and the third in ORF2. To conclude, the most prevalent form in ORF3 was P->L, which was stated as the fourth in ORF1.

HEV RdRp mutations previously reported in literature were also detected in S17, including K1383N and G1634R. In the case of G1634R, it already appeared in S7, when the second RBV 800mg treatment was started. Interestingly, an additional variant, D1384T, emerged at position 1384 in S7, and was maintained in the S17 consensus sequence.

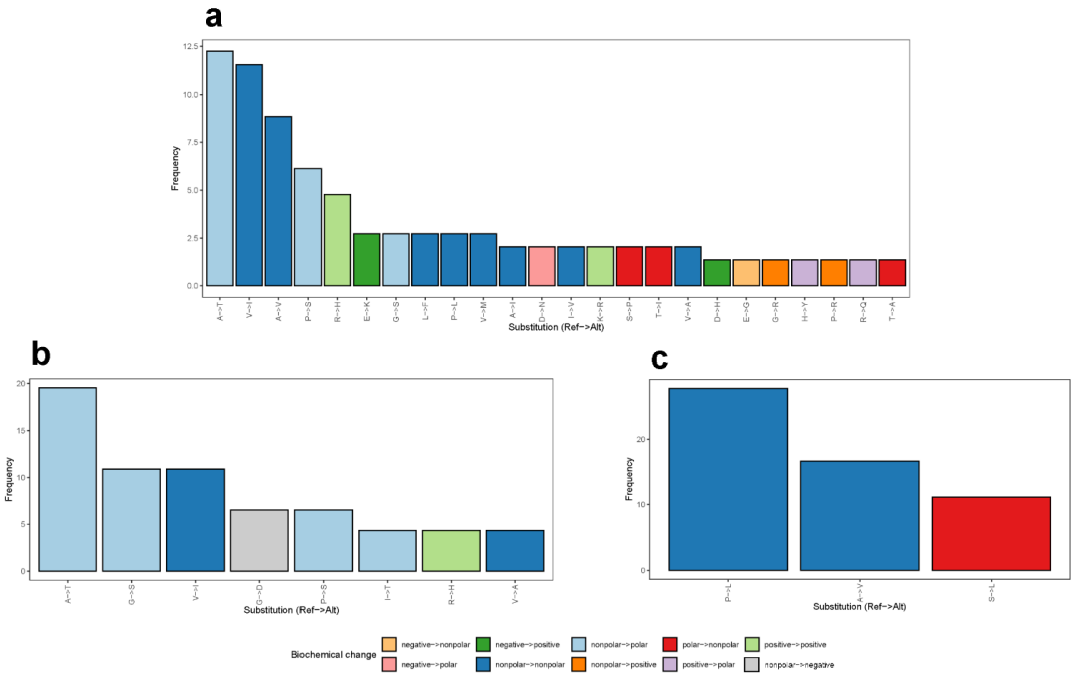


Figure 4. Histograms showing substitution frequency (as percentage) for ORF1 (a), ORF2 (b) and ORF3 (c) proteins for mutations present with n>1. Colours indicate the type of the biochemical change.

Discussion

In this study, we have analysed the evolution of HEV under sequential RBV treatment, revealing a set of amino acid substitutions that arose during prolonged antiviral pressure. HEV viral intra-host nucleotidic diversity was already addressed in our previous publications studying a single amplicon [12,21]. Nevertheless, amplicon sequencing strategy restricted the analysis to a region of the genome. The use of mNGS enabled full-genome coverage, albeit at the expense of haplotype resolution and sequencing depth.

Nucleotide variants presented a rising tendency from S1 to S17, highlighting the intrinsic capacity of the viral population to accumulate genomic diversity. The increase in variability may be attributed to the successive RBV treatment, as suggested by the concurrent rise in the Ts/Tv ratio[22,23]. Notably, the majority of variants that were maintained during the infection emerged only after the first treatment cessation and subsequent rebound, indicating that ribavirin likely enriched a broader mutant spectrum. This expanding mutant cloud may have facilitated the incorporation of higher fitness substitutions into the HEV quasispecies, contributing to posterior treatment failure[24], and potentially influencing future antiviral treatment.

Analysis of substitutions at the protein-level, underscore the importance of identifying variants that may critically influence patient response to RBV treatment. A marked contrast was observed between mutations detected at the consensus level and those present at intermediate allele frequencies (AF > 0.05 to AF > 0.5). This distinction is crucial for determining selective mutations within a highly variable viral population. Notably, in samples S7 and S17, the elevated overall mutational load may dilute the frequency of key substitutions, potentially masking variants that could impact treatment outcome.

To evaluate the potential impact of the substitutions observed during RBV treatment, we focused on the most frequently occurring amino acid changes and their biochemical properties. The most frequently change, A → T, represents a non-conservative substitution, shifting from a non-polar alanine to a polar threonine and thereby altering both polarity and local protein dynamics[25]. Other common substitutions include conservative changes (V → I, A → V, R → H), which generally maintain similar chemical properties and are more likely to be tolerated, as well as non-conservative or semi-conservative changes (G → S, P → L), which can influence protein flexibility, polarity or structural rigidity[26,27]. For example, the P → L substitution likely affects local rigidity due to loss of the proline-induced kink, while G → S introduces polarity. These observations

suggest that the A → T substitution, in particular, may play a role in HEV adaptation under antiviral pressure.

Among the previously described RBV resistance-associated mutations in the literature[17,28], only Y1587F was present at the consensus level in the last sample point. A variant affecting the same position as the previously reported D1384G, specifically D1384N, also emerged at consensus level in S17, resulting in the loss of the amino acid's negative charge. Interestingly, when examining the minor variants, a different substitution at the same position, D1384T, appeared, highlighting this site as a potential hotspot for variation associated to RBV resistance. At similar allele frequency, K1383N was also detectable. Of particular note, G1634R was present in S7 and maintained in S17. To assess whether additional mutations were present at low haplotype frequencies, we examined variants with AF < 0.05 and found that D1384N was already detectable in S7, while K1398R only appeared in S17 after three RBV treatments. Taken together, these findings suggest that a detailed analysis of minor variants could have predicted the emergence of key resistance to RBV, such as D1384N and G1634R, potentially allowing more informed treatment decisions.

To sum up, there is a critical importance for characterizing minor viral variants, including those present at the quasispecies level, when considering RBV monotherapy following treatment failure. Our findings reveal that consensus-level mutations, likely derived from pre-existing low-frequency variants, emerge under sequential RBV exposure together with viral replication, potentially contributing to antiviral resistance. These observations build upon our previous reports on HEV genomic adaptation during RBV treatment and highlight the limitations of relying solely on consensus sequence analyses. In light of the increasing recognition of RBV resistance in chronic HEV infection, our results reinforce the need for deep sequencing approaches to monitor intra-host viral diversity and support the development of alternative therapeutic strategies. In this study, we have analysed the evolution of HEV under sequential RBV treatment, revealing a set of amino acid substitutions that arose during prolonged antiviral pressure. Early detection of resistant mutations, even when present at low frequency, following the first treatment failure and after viral load recovery would have avoided an ineffective RBV monotherapy regimen with RBV. This could have prevented unnecessary drug exposure and further selection and fixation of resistance-associated mutations.

342 ***Methods***

343 ***Sample collection***

344 We report the clinical course of a 27-year-old patient who developed chronic HEV
345 infection after undergoing two kidney transplantations (Figure 1). The patient received
346 three different RBV treatment regimens during follow-up[18]. The plasma collection study
347 was approved by Vall d'Hebron University Hospital Ethics Committee, with the reference
348 numbers PR(AG)233-2018, PR(AG)118-2021, and PR(AG)287-2022.

349 The first course consisted of RBV 600 mg/day for three months. This regimen induced a
350 substantial decline in viral load, nearing elimination, but did not achieve HEV RNA
351 clearance. Renal function remained stable, and no signs of anaemia were observed. Given
352 the absence of virologic clearance, treatment was discontinued.

353 Several weeks later, the patient experienced virologic relapse, evidenced by a marked
354 rebound in HEV RNA levels and a concomitant increase in transaminases. A second course
355 of RBV was initiated at 800 mg/day, leading to a decline in HEV RNA with partial
356 biochemical improvement. By the fifth month, the dose was escalated to 1000 mg/day in
357 an attempt to further enhance the antiviral effect. Despite this, HEV RNA remained
358 detectable after two additional months of therapy, and treatment was stopped.

359 Ten days following discontinuation, a sharp virologic rebound was observed, with HEV
360 RNA levels increasing by three logs compared to end-of-treatment values.

361 ***RNA processing and sequencing***

362 HEV RNA was obtained by QIAamp MinElute Virus Spin extraction kit and QIAamp Viral
363 RNA Mini Kit (QIAGEN, Hilden, Germany) without the addition of the carrier. Following,
364 libraries were prepared using the TruSeq Stranded total RNA library kit (Illumina, San
365 Diego, CA, USA). Both extraction and library negative controls were included in the
366 experiment and sequenced.

367 Library quality was evaluated through the KAPA Library Quantification Kit (Roche Applied
368 Science, Pleasanton, CA, USA) in the Light Cycler 480 platform. Then, TapeStation 4200
369 system (Agilent, Santa Clara, CA, USA) was used for DNA profiling. At last, sequencing was
370 carried out in NextSeq2000 platform (Illumina, San Diego, CA, USA), with PhiX v3
371 (Illumina, San Diego, CA, USA) as internal control.

372 ***Bioinformatic analysis***

Sequencing data was retrieved from the sequencing platform in FASTQ format for all samples and negative controls. Then, files underwent an initial quality control (QC) assessment.

QC analysis began with the generation of quality reports using FastQC[29] and MultiQC[30]. Next, raw reads were trimmed using Trimmomatic[31] to remove adapter sequences, low-quality bases, and short reads. Duplicate sequences were removed using the clumpify tool from BBMAP suite[32]. To minimize the influence of technical contamination, BBDUk from BBMAP suite[32] was used to filter out reads present in the negative controls included in each sequencing run. Human reads were identified by mapping against the human reference genome (GRCh38) using Bowtie2[33]and subsequently removed from the dataset. The remaining non-human reads were subjected to an initial taxonomic classification using Kraken2[34] with the PlusPF database. Then, de-novo assembly of high-quality, non-human reads was performed using metaspades[35]. In parallel, reads were mapped to the complete genome of HEV (belonging to S1), through very sensitive local parameters with the use of bowtie2. Then, variants were identified through LoFreq[36] and consensus sequences were assembled through bcftools[37] applying variants whose allele frequency > 50%.

Opening reading frames (ORF) were identified through ORF finder (https://www.ncbi.nlm.nih.gov/orffinder/)[19].

All downstream analyses were conducted in R[38] (v4.2.5). Data processing and visualization such as graphs and statistical summaries were performed using tidyverse[39], ggplot2[40] and other relevant R packages.

Author Contribution

SC-C: methodology, investigation, formal analysis, visualization, tables and figures; writing original draft; M.I.-L.: software, formal analysis, data curation, and validation; writing original draft; CC: methodology, investigation, visualization, tables and figures; JG: conceptualization, methodology, software, formal data analysis, visualization; DG-C: methodology, investigation, and project administration; AR-S and RF: biochemical data, resources and methodology; SPR: methodology, investigation; MFC and DT: conceptualization, resources, and writing review and editing; JV-R: sample collection, clinical patient's follow-up; JC-RC, MR-B and MB: sample collection, clinical patient's follow-up, resources, formal analysis, and validation; JQ: conceptualization, methodology, validation, supervision, project administration, funding acquisition, and writing original draft. All authors have read and agreed to the published version of the manuscript.

Additional Information

Competing interest statements: The author(s) declare no competing interests.

Funding: This study was partially supported by Grants PI22/00258 and PI23/01065 funded by Instituto de Salud Carlos III (ISCIII) and cofounded by the European Union; Grant PID2021-126447OB-I00 funded by MCIN/AEI/ 10.13039/501100011033 and by ERDF A way of making Europe; Consorcio Centro de Investigación Biomédica en Red (CIBEREHD). C.C. received a predoctoral fellowship from the Vall d'Hebron Institute of Research (VHIR). M.I.-L. was supported by a fellowship from the "La Caixa" Foundation (ID 100010434), with code "LCF/BQ/DR23/12000020". S.C-C. received support from the Spanish Ministry of Education grant FPU21/04150.

Data Availability Statement: Sequencing data that support the findings of this study will be openly available in the GenBank Sequence Read Archive database after publication (submitted as SUB15794654).

Code Availability Statement: All scripts used for data analysis will be made available upon request by contacting the corresponding author via email.

Inclusion & Ethics statement: In this study we have used leftover plasma or serum samples from diagnosis procedures of anonymized patients infected by one of the viruses described in the manuscript. The study was approved by Vall d'Hebron University Hospital Ethics Committee, with the reference numbers PR(AG)233-2018, PR(AG)118-2021, and PR(AG)287-2022.

Informed Consent Statement: Informed consent was obtained from the subject involved in the study.

Acknowledgments: We would like to acknowledge Roche Diagnostics S.L.U., who partially support the cost of publishing this manuscript.

References

1. Hepatitis E. https://www.who.int/news-room/fact-sheets/detail/hepatitis-e. Accessed on 25/11/2025.
2. Songtanin, B., Molehin, A. J., Brittan, K., Manatsathit, W. & Nugent, K. Hepatitis E Virus Infections: Epidemiology, Genetic Diversity, and Clinical Considerations. *Viruses 2023, Vol. 15, Page 1389* **15**, 1389 (2023).
3. Khuroo, M. S., Khuroo, M. S. & Khuroo, N. S. Hepatitis E: Discovery, global impact, control and cure. *World J Gastroenterol* **22**, 7030–7045 (2016).

4. Riveiro-Barciela, M. *et al.* Phylogenetic demonstration of hepatitis E infection transmitted by pork meat ingestion. *J Clin Gastroenterol* **49**, 165–168 (2015).

5. Boxman, I. L. A. *et al.* Monitoring of pork liver and meat products on the Dutch market for the presence of HEV RNA. *Int J Food Microbiol* **296**, 58–64 (2019).

6. Bernal, W., Auzinger, G., Dhawan, A. & Wendon, J. Acute liver failure. *The Lancet* **376**, 190–201 (2010).

7. Luo, Q. *et al.* Viral hepatitis E: Clinical manifestations, treatment, and prevention. *Liver Res* **8**, 11–21 (2024).

8. Jha, A. K., Kumar, G., Dayal, V. M., Ranjan, A. & Suchismita, A. Neurological manifestations of hepatitis E virus infection: An overview. *World J Gastroenterol* **27**, 2090–2104 (2021).

9. Kovvuru, K. *et al.* Renal manifestations of hepatitis E among immunocompetent and solid organ transplant recipients. *World J Hepatol* **14**, 516–524 (2022).

10. Sridhar, S. *et al.* Donor-Derived Genotype 4 Hepatitis E Virus Infection, Hong Kong, China, 2018 - Volume 25, Number 3—March 2019 - Emerging Infectious Diseases journal - CDC. *Emerg Infect Dis* **25**, 425–433 (2019).

11. Dalton, H. R. *et al.* EASL Clinical Practice Guidelines on hepatitis E virus infection. *J Hepatol* **68**, 1256–1271 (2018).

12. Colomer-Castell, S. *et al.* In-Host HEV Quasispecies Evolution Shows the Limits of Mutagenic Antiviral Treatments. *Int J Mol Sci* **24**, 17185 (2023).

13. Lampejo, T. Sofosbuvir in the Treatment of Hepatitis E virus Infection: A Review of in vitro and in vivo Evidence. *J Clin Exp Hepatol* **12**, 1225–1237 (2022).

14. Domingo, E. & Holland, J. J. Mutation rates and rapid evolution of RNA viruses. in *Evolutionary biology of viruses* (ed. Morse, S. S.) 161–184 (Raven Press, New York, 1994).

15. Todt, D., Walter, S., Brown, R. J. P. & Steinmann, E. Mutagenic Effects of Ribavirin on Hepatitis E Virus—Viral Extinction versus Selection of Fitness-Enhancing Mutations. *Viruses 2016, Vol. 8, Page 283* **8**, 283 (2016).

16. Muñoz-Chimeno, M., Rodriguez-Paredes, V., García-Lugo, M. A. & Avellon, A. Hepatitis E genotype 3 genome: A comprehensive analysis of entropy, motif conservation, relevant mutations, and clade-associated polymorphisms. *Front Microbiol* **13**, 1011662 (2022).

17. van Tong, H. *et al.* Hepatitis E Virus Mutations: Functional and Clinical Relevance. *EBioMedicine* **11**, 31–42 (2016).

18. Gregori, J. *et al.* Quasispecies Fitness Partition to Characterize the Molecular Status of a Viral Population. Negative Effect of Early Ribavirin Discontinuation in a Chronically Infected HEV Patient. *Int J Mol Sci* **23**, 14654 (2022).

19. ORFfinder Home - NCBI. https://www.ncbi.nlm.nih.gov/orffinder/. Accessed on 25/11/2025.

20. Koonin, E. V. *et al.* Computer-assisted assignment of functional domains in the nonstructural polyprotein of hepatitis E virus: Delineation of an additional group of positive-strand RNA plant and animal viruses. *Proc Natl Acad Sci U S A* **89**, 8259–8263 (1992).

21. Gregori, J. *et al.* Quasispecies Fitness Partition to Characterize the Molecular Status of a Viral Population. Negative Effect of Early Ribavirin Discontinuation in a Chronically Infected HEV Patient. *Int J Mol Sci* **23**, 14654 (2022).

22. Aljabr, W. *et al.* Investigating the Influence of Ribavirin on Human Respiratory Syncytial Virus RNA Synthesis by Using a High-Resolution Transcriptome Sequencing Approach. *J Virol* **90**, 4876–4888 (2016).

23. Lyons, D. M. & Lauring, A. S. Evidence for the Selective Basis of Transition-to-Transversion Substitution Bias in Two RNA Viruses. *Mol Biol Evol* **34**, 3205–3215 (2017).

24. Todt, D. *et al.* In vivo evidence for ribavirin-induced mutagenesis of the hepatitis E virus genome. *Gut* **65**, 1733 (2016).

25. Podoly, E., Hanin, G. & Soreq, H. Alanine-to-threonine substitutions and amyloid diseases: Butyrylcholinesterase as a case study. *Chem Biol Interact* **187**, 64–71 (2010).

26. Pakula, A. A. & Sauer, R. T. Genetic analysis of protein stability and function. *Annu Rev Genet* **23**, 289–310 (1989).

27. Zito, F., Kuras, R., Choquet, Y., Kössel, H. & Wollman, F. A. Mutations of cytochrome b6 in Chlamydomonas reinhardtii disclose the functional significance for a proline to leucine conversion by petB editing in maize and tobacco. *Plant Mol Biol* **33**, 79–86 (1997).

28. Gömer, A. *et al.* Emergence of resistance-associated variants during sofosbuvir treatment in chronically infected hepatitis E patients. *Hepatology* **78**, 1882 (2023).

29. Babraham Bioinformatics - FastQC A Quality Control tool for High Throughput Sequence Data. https://www.bioinformatics.babraham.ac.uk/projects/fastqc/.

30. Ewels, P., Magnusson, M., Lundin, S. & Käller, M. MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* **32**, 3047 (2016).

31. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).

32. Bushnell, A. *et al.* Mapping single molecule sequencing reads using basic local alignment with successive refinement (BLASR): Application and theory. *BMC Bioinformatics* **13**, (2014).

33. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**, 357 (2012).

34. Wood, D. E., Lu, J. & Langmead, B. Improved metagenomic analysis with Kraken 2. *Genome Biology 2019 20:1* **20**, 257- (2019).

35. Bankevich, A. *et al.* SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. *Journal of Computational Biology* **19**, 455 (2012).

518 36. Wilm, A. *et al.* LoFreq: a sequence-quality aware, ultra-sensitive variant caller for
519 uncovering cell-population heterogeneity from high-throughput sequencing datasets.
520 *Nucleic Acids Res* **40**, 11189–11201 (2012).

521 37. Danecek, P. *et al.* Twelve years of SAMtools and BCFtools. *Gigascience* **10**, giab008
522 (2021).

523 38. R Core Team (2022) R A Language and Environment for Statistical Computing. R
524 Foundation for Statistical Computing, Vienna. - References - Scientific Research
525 Publishing. https://www.scirp.org/reference/referencespapers?referenceid=3456808.

526 39. Wickham, H. *et al.* Welcome to the Tidyverse. *J Open Source Softw* **4**, 1686 (2019).

527 40. Wickham, H. *Ggplot2: Elegant Graphics for Data Analysis*. (Springer-Verlag New York,
528 2016).

529

530 **Figure and Table Legends**

531 **Figure 1.** Infection course diagram showing the extraction date of the sample and its
532 corresponding VL in logarithmic scale. Vertical dotted lines indicate the RBV treatment
533 application (red) or End of treatment (Blue). Samples (S1, S2, S6, S7 and S17) with
534 complete genome coverage are highlighted.

535 **Figure 2.** Lollipop Chart showing the allele frequency (Y axis) from all the variants
536 identified in S1, S2, S6, S7 and S17. Variants are distributed following the genome position
537 (X axis). All variants were filtered from 0.05 AF and 50 reads.

538 **Figure 3.** Line point diagram showing the nucleotide mutations that were present in three
539 study points or more. Y axis corresponds to mutation allele frequency while X axis shows
540 the samples where the mutations are present. Colours represent the sample in which the
541 mutation was first detected.

542 **Figure 4.** Histograms showing substitution frequency (as percentage) for ORF1 (a), ORF2
543 (b) and ORF3 (c) proteins for mutations present with n>1. Colours indicate the type of the
544 biochemical change.

545 **Table 1.** VL in Log (VL) and three different sequencing parameters (CPM, Mean depth and
546 % Covered) are shown for all samples. CPM accounts for counts per millions; % Covered
547 constitutes the percentage of HEV genome covered by sequencing.

548 **Table 2.** Mutations occurring at the consensus level from S2, S6, S7 and S17 in pORF1,
549 pORF2 and pORF3. Its presence between consensus threshold and AF < 0.05 is showed in
550 the right column.

551

5.1.3 Study 3: Supplementary Material

Whole-genome Analysis of HEV Under Sequential Ribavirin Pressure Reveals Early Minor Variants Predicting Resistance.

Whole-genome Analysis of HEV Under Sequential Ribavirin Pressure Reveals Early Minor Variants Predicting Resistance.

Sergi Colomer-Castell[1,2,3], Marta Ibañez-Lligoña[1,3,4], Carolina Campos[1,2,3], Josep Gregori[1], Damir Garcia-Cehic[1,2,5], Ariadna Rando-Segura[3,6], Roser Ferrer[4,7], Sofía P. Ruzo[1], Maria Francesca Cortese[1,3,6], David Tabernero[1,3,6], Judit Vico-Romero[1], Juan Carlos Ruiz-Cobo[1,3,4], Maria Buti[1,3,4], Mar Riveiro-Barciela[1,3], Josep Quer[1,2,3,4,*].

[1]Liver Diseases-Viral Hepatitis, Liver Unit, Vall d'Hebron Institut de Recerca (VHIR), Vall d'Hebron Hospital Universitari, Vall d'Hebron Barcelona Hospital Campus, Barcelona.

[2]Biochemistry and Molecular Biology Department, Universitat Autònoma de Barcelona (UAB), Bellaterra.

[3]Centro de Investigación Biomédica en Red de Enfermedades Hepáticas y Digestivas (CIBERehd), Madrid.

[4]Medicine Department, Universitat Autònoma de Barcelona (UAB), Bellaterra.

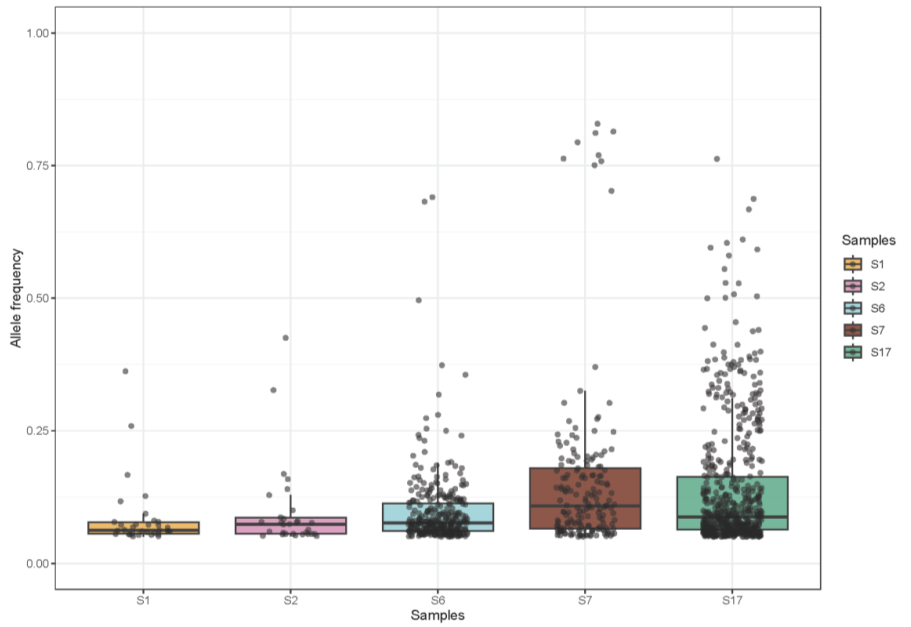
[5]Department of Radiation Oncology, University of Cincinnati, College of Medicine, Cincinnati, Ohio, USA

[6]Microbiology Department, Vall d'Hebron Institut de Recerca (VHIR), Vall d'Hebron Hospital Universitari, Vall d'Hebron Barcelona Hospital Campus, Universitat Autònoma de Barcelona, Barcelona.

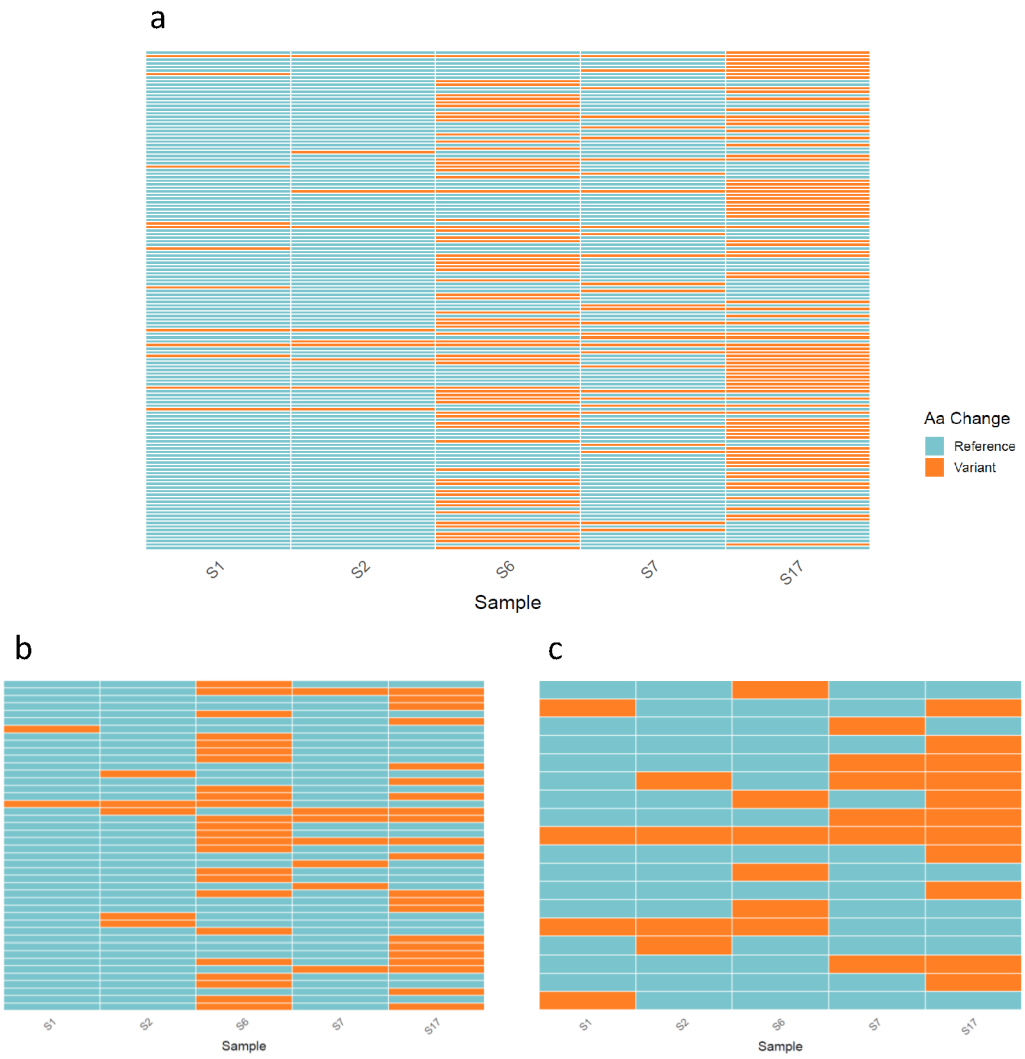
[7]Biochemistry Department, Vall d'Hebron Institut de Recerca (VHIR), Vall d'Hebron Hospital Universitari, Vall d'Hebron Barcelona Hospital Campus, Barcelona.

***Corresponding author:**

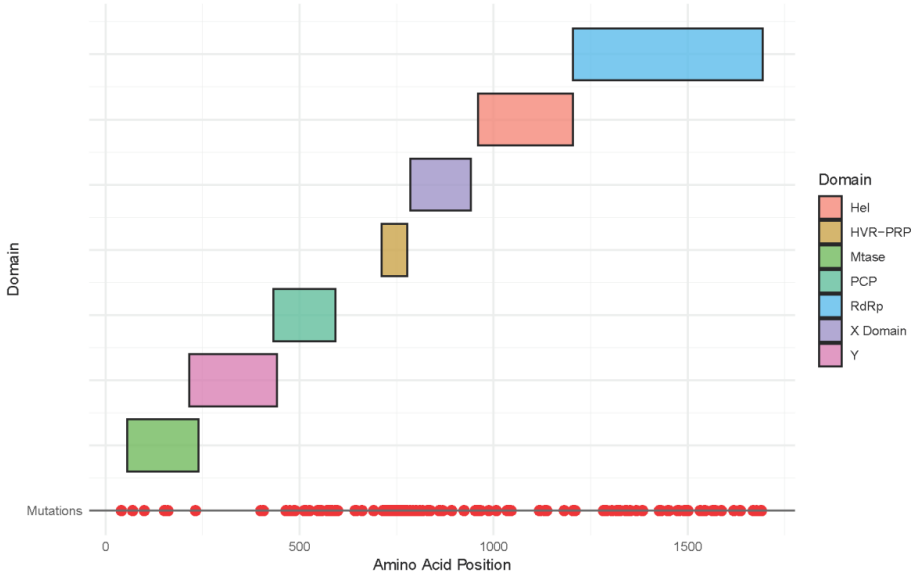
Josep Quer josep.quer@vhir.org



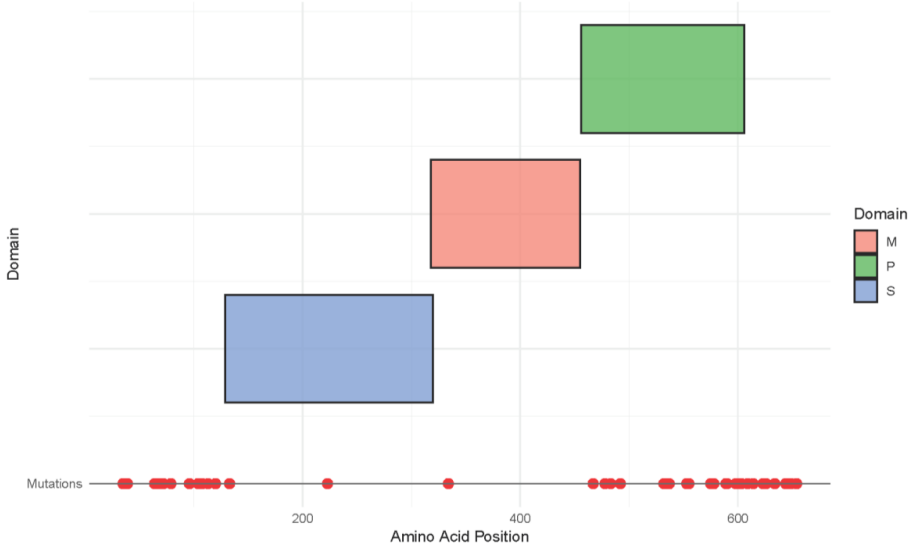
Suppl. Figure S1. Allele frequency distribution per sample presented in a Box-plot diagram.



Suppl. Figure S2. Heatmap depicting the presence (orange) or absence (blue) of substitutions with an AF > 0.05. The Y axis lists each identified substitution, and the X axis represents the analysed samples. Panels correspond to the different HEV proteins: (a) pORF1, (b) pORF2 and (c) pORF3.



Suppl. Figure S3. Domain distribution of substitutions in HEV pORF1 (AF > 0.05). Domains are showed in coloured boxes vertically across the Y axis. Substitutions are represented by red dots in the X axis line.



Suppl. Figure S4. Domain distribution of substitutions in HEV pORF2 (AF > 0.05). Domains are showed in coloured boxes vertically across the Y axis. Substitutions are represented by red dots in the X axis line.

Suppl. Table S1. Substitutions found with AF > 0.05 in pORF1 for S1, S2, S6, S7 and S17. nVariants account for the number of samples carrying the substitution in the quasispecies.

Mutation	S1	S2	S6	S7	S17	nVariant
V41M	Reference	Reference	Variant	Reference	Reference	1
V70T	Reference	Reference	Reference	Reference	Variant	1
P100S	Reference	Reference	Variant	Reference	Reference	1
R152H	Reference	Reference	Reference	Reference	Variant	1
S154P	Reference	Reference	Reference	Variant	Reference	1
A161V	Reference	Reference	Reference	Reference	Variant	1
D232N	Reference	Reference	Variant	Variant	Variant	3
V400I	Reference	Reference	Variant	Reference	Variant	2
I407V	Reference	Reference	Variant	Reference	Reference	1
K465R	Reference	Reference	Variant	Reference	Reference	1
A466T	Reference	Reference	Variant	Reference	Reference	1
K475R	Reference	Reference	Reference	Variant	Variant	2
E486K	Reference	Reference	Variant	Variant	Variant	3
A488T	Reference	Reference	Reference	Reference	Variant	1
P512L	Reference	Reference	Variant	Reference	Reference	1
A513T	Reference	Reference	Variant	Reference	Reference	1
V517I	Reference	Reference	Reference	Reference	Variant	1
R526H	Reference	Reference	Reference	Reference	Variant	1
R547C	Reference	Reference	Reference	Reference	Variant	1
T549A	Reference	Reference	Reference	Reference	Variant	1
V555A/I	Reference	Reference	Reference	Variant	Variant	2
G557S	Reference	Reference	Variant	Reference	Variant	2
T571I	Reference	Reference	Variant	Reference	Reference	1
V577I	Reference	Reference	Variant	Reference	Reference	1
A580V	Reference	Reference	Variant	Reference	Reference	1
A584T	Reference	Reference	Reference	Reference	Variant	1
V591I	Reference	Reference	Variant	Reference	Reference	1
A596V	Reference	Reference	Variant	Reference	Reference	1
R598H	Reference	Reference	Reference	Reference	Variant	1
G643E	Reference	Variant	Variant	Reference	Variant	3
A647T	Reference	Reference	Variant	Reference	Reference	1
R661H	Reference	Variant	Variant	Variant	Variant	4
A691T	Reference	Reference	Reference	Reference	Variant	1
S715F	Reference	Reference	Variant	Reference	Reference	1
P717L	Reference	Reference	Variant	Reference	Reference	1
A720T	Reference	Reference	Variant	Reference	Variant	2
V721A/I	Variant	Reference	Reference	Reference	Variant	2
A722V	Reference	Reference	Variant	Reference	Reference	1
A723I	Reference	Reference	Reference	Reference	Variant	1
P724S	Reference	Reference	Variant	Variant	Variant	3
A725V	Reference	Reference	Reference	Reference	Variant	1
A726V	Reference	Reference	Variant	Reference	Reference	1
R731hev3Y	Reference	Reference	Reference	Reference	Variant	1
P733L	Reference	Reference	Reference	Reference	Variant	1
I737T	Reference	Reference	Variant	Variant	Variant	3
W738R	Variant	Variant	Variant	Variant	Variant	5
V739M	Reference	Reference	Reference	Variant	Variant	2
P741R	Variant	Reference	Reference	Reference	Reference	1
S744P	Variant	Reference	Variant	Reference	Reference	2

E749hev3K	Reference	Reference	Variant	Reference	Reference	1
V752hev3I	Reference	Reference	Reference	Reference	Variant	1
A748I	Reference	Reference	Reference	Reference	Variant	1
A750T	Reference	Reference	Reference	Reference	Variant	1
E755G	Reference	Reference	Variant	Variant	Variant	3
A757V	Reference	Reference	Reference	Reference	Variant	1
P759S	Reference	Reference	Reference	Reference	Variant	1
S761R	Reference	Reference	Variant	Reference	Reference	1
P762S	Reference	Reference	Variant	Reference	Variant	2
A763T	Reference	Reference	Reference	Reference	Variant	1
A767T/I	Reference	Reference	Reference	Variant	Variant	2
R777hev3H	Reference	Reference	Reference	Reference	Variant	1
K778hev3R	Reference	Reference	Reference	Variant	Reference	1
P770S	Reference	Reference	Variant	Reference	Reference	1
S774P/*	Reference	Reference	Variant	Variant	Variant	3
P775R	Reference	Reference	Reference	Variant	Reference	1
P785S	Reference	Reference	Variant	Reference	Reference	1
A792V	Reference	Reference	Reference	Reference	Variant	1
D799H	Variant	Variant	Reference	Reference	Reference	2
D801N	Reference	Reference	Reference	Variant	Variant	2
G810C	Variant	Reference	Variant	Reference	Variant	3
L817I	Variant	Reference	Reference	Reference	Reference	1
H830Y	Variant	Variant	Variant	Variant	Variant	5
P831S/W	Variant	Variant	Variant	Variant	Variant	5
I835L	Reference	Reference	Variant	Reference	Reference	1
R837H	Reference	Reference	Reference	Reference	Variant	1
V860I	Reference	Reference	Reference	Reference	Variant	1
E861G	Variant	Variant	Variant	Reference	Variant	4
H862R	Reference	Variant	Variant	Reference	Reference	2
A869T	Reference	Reference	Reference	Variant	Reference	1
Q892R	Variant	Reference	Reference	Reference	Reference	1
A924V	Reference	Reference	Variant	Reference	Reference	1
E953K	Reference	Reference	Reference	Reference	Variant	1
A959T	Reference	Reference	Reference	Reference	Variant	1
C961Y	Reference	Reference	Reference	Reference	Variant	1
A962T	Reference	Reference	Reference	Reference	Variant	1
V967I	Reference	Reference	Reference	Reference	Variant	1
G988R/K	Reference	Reference	Reference	Variant	Variant	2
R1007*	Reference	Reference	Variant	Reference	Reference	1
P1035S	Reference	Reference	Reference	Reference	Variant	1
L1041F	Reference	Reference	Reference	Reference	Variant	1
R1045Q	Reference	Reference	Reference	Reference	Variant	1
P1118S	Reference	Reference	Reference	Reference	Variant	1
A1119V	Reference	Reference	Variant	Reference	Reference	1
T1120I	Reference	Reference	Reference	Reference	Variant	1
A1132V	Reference	Reference	Reference	Reference	Variant	1
A1137V	Reference	Reference	Variant	Reference	Reference	1
V1182I	Reference	Reference	Reference	Variant	Variant	2
L1205F	Reference	Reference	Variant	Reference	Reference	1
V1210M	Reference	Reference	Variant	Reference	Reference	1

V1283M	Reference	Reference	Reference	Reference	Variant	1
V1285I	Reference	Reference	Reference	Variant	Reference	1
D1290N	Reference	Reference	Variant	Variant	Variant	3
V1292I	Reference	Reference	Reference	Reference	Variant	1
V1305I	Reference	Reference	Variant	Reference	Variant	2
T1317A	Reference	Variant	Reference	Reference	Reference	1
A1323T	Reference	Reference	Variant	Reference	Reference	1
D1326G	Reference	Reference	Variant	Reference	Variant	2
P1340L	Reference	Reference	Variant	Reference	Reference	1
A1343T	Reference	Reference	Variant	Reference	Reference	1
V1352I	Reference	Reference	Variant	Variant	Variant	3
L1368F	Reference	Reference	Variant	Reference	Reference	1
K1383N	Reference	Reference	Reference	Reference	Variant	1
D1384T	Reference	Reference	Reference	Reference	Variant	1
A1427T	Reference	Reference	Reference	Variant	Reference	1
L1429F	Reference	Reference	Reference	Variant	Reference	1
P1431*	Reference	Reference	Variant	Reference	Reference	1
S1449L	Reference	Reference	Variant	Reference	Reference	1
G1452S	Reference	Reference	Reference	Reference	Variant	1
G1474S	Reference	Reference	Reference	Reference	Variant	1
V1478A	Reference	Reference	Variant	Reference	Reference	1
R1491Q	Reference	Reference	Reference	Reference	Variant	1
V1496I	Reference	Reference	Reference	Reference	Variant	1
A1499V	Reference	Reference	Variant	Reference	Reference	1
I1501V	Reference	Reference	Reference	Variant	Variant	2
A1532T	Reference	Reference	Variant	Variant	Reference	2
I1533V/A	Reference	Reference	Variant	Variant	Variant	3
F1543L	Reference	Reference	Reference	Reference	Variant	1
V1545I	Reference	Reference	Variant	Reference	Reference	1
R1564H	Reference	Reference	Reference	Reference	Variant	1
N1565S	Reference	Reference	Reference	Variant	Reference	1
A1568T	Reference	Reference	Reference	Reference	Variant	1
G1572S	Reference	Reference	Reference	Reference	Variant	1
Y1587F	Reference	Reference	Reference	Reference	Variant	1
E1619K	Reference	Reference	Variant	Reference	Reference	1
G1634R	Reference	Reference	Reference	Variant	Variant	2
T1636I	Reference	Reference	Variant	Reference	Reference	1
D1669H	Reference	Reference	Variant	Reference	Reference	1
V1678I	Reference	Reference	Variant	Reference	Reference	1
I1688I	Variant	Variant	Reference	Reference	Reference	2
H1690Y	Reference	Reference	Reference	Reference	Variant	1

Suppl. Table S2. Substitutions found with AF > 0.05 in pORF2 for S1, S2, S6, S7 and S17. nVariants account for the number of samples carrying the substitution in the quasispecies.

Mutation	S1	S2	S6	S7	S17	nVariant
G35S	Reference	Reference	Reference	Reference	Variant	1
G36D/S	Reference	Reference	Variant	Reference	Variant	2
G39S	Reference	Reference	Reference	Variant	Reference	1
A64T	Reference	Reference	Reference	Reference	Variant	1
V66I	Reference	Reference	Reference	Reference	Variant	1
V67I	Reference	Reference	Reference	Reference	Variant	1
P68S	Reference	Variant	Reference	Variant	Variant	3
A72T	Reference	Reference	Reference	Reference	Variant	1
P79L	Variant	Variant	Variant	Reference	Reference	3
A96T	Reference	Reference	Reference	Reference	Variant	1
P104S	Reference	Reference	Variant	Reference	Reference	1
A105T/I	Reference	Reference	Variant	Reference	Variant	2
A108T	Reference	Reference	Variant	Reference	Reference	1
I113T	Reference	Reference	Reference	Reference	Variant	1
V120I	Reference	Reference	Variant	Reference	Reference	1
R133Q	Reference	Reference	Variant	Reference	Reference	1
D223G	Reference	Reference	Variant	Reference	Reference	1
G334S	Reference	Reference	Reference	Reference	Variant	1
A467T	Reference	Reference	Reference	Reference	Variant	1
A478T	Reference	Reference	Variant	Reference	Reference	1
T483A	Reference	Reference	Reference	Reference	Variant	1
M492V	Reference	Reference	Variant	Reference	Reference	1
Y532D	Reference	Reference	Variant	Variant	Variant	3
T535S	Reference	Reference	Variant	Reference	Reference	1
Y537F	Reference	Reference	Variant	Reference	Reference	1
T553I	Reference	Reference	Variant	Reference	Reference	1
A555T	Reference	Reference	Variant	Reference	Reference	1
A575T	Reference	Reference	Reference	Variant	Variant	2
R578H	Reference	Reference	Reference	Reference	Variant	1
G589S	Reference	Reference	Variant	Reference	Reference	1
A590V	Reference	Reference	Variant	Reference	Variant	2
V598I	Reference	Reference	Variant	Reference	Reference	1
G599D	Reference	Reference	Variant	Reference	Reference	1
V600A	Variant	Reference	Reference	Reference	Reference	1
P603S	Reference	Reference	Variant	Variant	Variant	3
V609I	Reference	Reference	Reference	Reference	Variant	1
I614T	Reference	Reference	Variant	Reference	Reference	1
F623L	Reference	Variant	Reference	Reference	Reference	1
F626S	Reference	Variant	Reference	Reference	Reference	1
G634D	Reference	Reference	Reference	Variant	Reference	1
I644V	Reference	Reference	Variant	Variant	Variant	3
Q648E	Reference	Reference	Variant	Reference	Variant	2
R649H	Reference	Variant	Reference	Reference	Reference	1
V654A	Reference	Reference	Variant	Reference	Reference	1

Suppl. Table S3. Substitutions found with AF > 0.05 in pORF3 for S1, S2, S6, S7 and S17. nVariants account for the number of samples carrying the substitution in the quasispecies.

Mutation	S1	S2	S6	S7	S17	nVariant
R25Q	Reference	Reference	Reference	Variant	Variant	2
A27V	Variant	Reference	Reference	Reference	Reference	1
V40M	Reference	Reference	Variant	Reference	Reference	1
S55L	Reference	Variant	Reference	Variant	Variant	3
T69M	Reference	Reference	Reference	Variant	Reference	1
R72H	Reference	Reference	Variant	Reference	Variant	2
A79V	Reference	Reference	Reference	Reference	Variant	1
D81G	Reference	Variant	Reference	Reference	Reference	1
S82N	Reference	Reference	Reference	Variant	Variant	2
H83Y	Variant	Variant	Variant	Reference	Reference	3
P84L	Reference	Reference	Reference	Reference	Variant	1
A85V	Reference	Reference	Reference	Variant	Variant	2
P86L	Reference	Reference	Variant	Reference	Reference	1
S87L	Reference	Reference	Reference	Reference	Variant	1
P89L	Reference	Reference	Reference	Reference	Variant	1
P90L	Variant	Variant	Variant	Variant	Variant	5
T93I	Variant	Reference	Reference	Reference	Variant	2
P101L	Reference	Reference	Variant	Reference	Reference	1

6. DISCUSSION

DISCUSSION

6.1 A Novel Strategy for the Analysis of Viral Quasispecies

The transition from clonal to next-generation sequencing (NGS) methods to study diversity and complexity of highly variable genomes, such as RNA viruses, has yielded to advanced diversity analysis tools, based in incidence, abundance, and functionality. In previous studies using clonal sequences, it was established that at sequencing depths of ≥10,000 reads per sample, haplotypes present above 0.5% abundance could be identified without erroneous mutations and with high confidence. Following this approach detection of resistant mutants as minor variants at ~1% frequency was reliable when used for diagnosis, and also provide accurate estimates of quasispecies complexity[168].Following this conservative cut-off, we were conscient that the analysis inevitably led to the loss of a substantial fraction of real viral variants present below 0.5%, which nonetheless belonged to the true viral population. This limitation was particularly evident when treating patients with a mutagen, such as Ribavirin (RBV). RBV is a mutagen that drives the production of low-fitness viral quasispecies, effectively eliminating the viral population through accumulated genetic damage rather than direct polymerase inhibition, characterized by G-to-A and C-to-U nucleotide transitions[119]. The consequence for the viral population when subjected to antiviral treatment using RBV or any other drug with mutagenic properties, is that a huge number of haplotypes are expected to fall below this abundance threshold. To overcome this constraint and enable the analysis of low-frequency but biologically meaningful variants, new analytical strategies are required.

Herein we have showed three different studies that cover the HEV quasispecies with different but related objectives highlighting the importance of monitoring viral quasispecies using NGS, particularly during chronic infection courses.

In the first study, we have developed and tested the new quasispecies fitness fraction partition (QFF) method to characterize the molecular status of a viral population by dividing haplotypes into four distinct categories based on their frequency. The frequency might serve as an indirect measurement for the virus fitness. QFF reflects the biological behaviour of the population instead of summarizing the quasispecies structure into a single numerical value. Between the four fractions, the master haplotype represents the currently domain strain. Emerging haplotypes, with frequencies higher than 1%, have the potential to compete and eventually replace the master strain, indicating how the environment is selecting the population. The previously defined rare haplotype load (RHL)[165], with frequencies here defined between 1% and 0.1%, and finally, a fourth category composed by the very rare haplotypes with frequencies lower than 0.1% belonging to haplotypes with very low fitness including defective genomes. The representation of each fraction using a different colour in histograms is highly valuable for visualizing the evolution of the quasispecies within a host, while the combination with phylogenetic analysis helps to understand the real infection dynamics.

To show the QFF performance in different viral distribution systems we have applied it to distinct data from amplicon sequencing experiments already performed in the laboratory. We have selected three RNA viral populations with different diversity structure, a viral clone from a virus showing low diversity since this virus uses an RNA-dependent RNA polymerase with proofreading mechanisms (SARS-CoV-2), a viral population of HCV in cell culture subjected to consecutive viral passages and treated in short-time passages to different antiviral drugs showing moderate variability, and a naturally HEV chronically infected patient treated with three dosage of RBV showing an extremely high variable viral population at the nucleotide level.

First, the QFF applied to a technical SARS-Cov-2 clone allowed us to confirm the reproducibility of the analysis and the technical limits, from both the clone synthesis and laboratory PCRs and sequencing. Within this case, the

longitudinal analysis in the same batch allow to compare between different samples in the same experiment.

Secondly, we evaluated how the QFF varied in a controlled *in vitro* experiment involving HCV-infected cells and treatments with RBV and favipiravir (mutagenic agents), and Sofosbuvir (Polymerase inhibitor). The results showed that the samples under mutagenic treatment showed an increase in the non-master haplotypes in respect to the control. Of note, there is an important frequency of emergent haplotypes, while both RHL 1 and 0.1 represent approximately half of the population haplotypes, possibly due to the mechanism of action of the drugs applied, which favour the variability inside the quasispecies. In the case of Sofosbuvir, the master shows a higher frequency than the control, which might reinforce the paper of sofosbuvir in reducing the replication rate and diminishing the formation of new minor variants.

Finally, the monitorization of the frequencies through QFF has been shown useful to identify the response of the viral population to the RBV treatments in a first HEV patient that underwent three different regimes, where the relapse was treated with a higher RBV dose and the development of a resistance could be exhibited. The important presence of haplotypes inside the RHL even at the first steps of the infection reveals how unstructured the HEV quasispecies is. These effects got even amplified at some points of the viral infection, especially after the third and last treatment of RBV with a dose of 1000 mg/day. In parallel, emergent haplotypes with an enhanced fitness also increased its frequency. However, there is partial restoration of the master at some points, coinciding with the lowest viral load points.

To complement the HEV course infection study, we compared the nucleotide QFF (genotype level) with the protein QFF (phenotype level) profiles. Here we shown that the viral population is able to control the variability through the generation of synonymous mutations, leading to a very conserved master

sequence that dominates the quasispecies with frequencies higher than 50%. However, when the higher RBV dose is applied, there is an important generation of emergent haplotypes.

Taken together, QFF has been shown to be a valuable tool for the study of highly variable systems such as viral quasispecies. It provides a highly visual and intuitive analytical framework that complements the Hill Number Profile (HNP) and other variability indices described in literature. Nevertheless, QFF is not exempt of limitations, since it relies on amplicon-based sequencing combined with very high sequencing depth to accurately uncover the Rare Haplotype Load (RHL).

6.2 HEV Quasispecies Dynamics during Ribavirin Treatment: Genotypic, Phenotypic, and Clinical Insights

Both, first and second studies here included, covered in a deep manner the effect of RBV treatment on two patients who underwent three drug regimes. It should be noted that our findings and conclusions are limited due to the low number of eligible patients. Actually, access to samples from chronically HEV-infected patient is limited, obtaining sequential samples is even more challenging, and having longitudinal samples from two patients followed over several years is rare. In our case the identification of two such patients have represented a valuable opportunity to uncover our hypothesis.

In both cases, the viral population was predominantly composed of RHL haplotypes, representing more than half of the quasispecies. In the last four points of monitoring, emergent haplotypes also accounted for a substantial fraction of the population, while RHL0.1 remained dominant. Interestingly, in both cases we observed a profound discrepancy between nucleotide and amino acid QFF patterns: the population was extremely variable at the nucleotide level, with even the master sequences present at frequencies

lower than 10%, yet the corresponding master phenotype accounted for approximately 60% at the amino acid level. These results suggest that, while RBV pressure generates an unstructured collection of genomes at the nucleotide level, the phenotypic structure remains remarkably stable. This stability likely reflects the generation of a rich set of synonymous haplotypes, enhanced by the high viral load, that preserve the same functional master phenotype, effectively masking functional stability despite high genetic diversity.

According to quasispecies theory, in environments with high mutation rates, such as those accelerated by RBV treatment, viral populations occupying regions with higher density of neutral mutations may experience an advantageous evolutionary landscape. Consequently, migration of the viral quasispecies toward these “flatter” regions of the fitness landscape allows the virus to maintain its functional master phenotype despite the accumulation of numerous mutations. This expansion, in combination with higher viral loads might be the explanation of how the virus overcome the error catastrophe via lethal mutagenesis, converting the RBV as an evolutionary catalyst instead of achieving viral extinction.

In RNA viruses replicating at high viral loads using an RNA-dependent RNA polymerase lacking of proofreading activity, monotherapy treatments often lead to failure. The use of mutagens to increase the mutation rate, potentially driving the virus beyond the error threshold and inducing error catastrophe, has been proposed by Esteban Domingo’s team as a strategy for lethal mutagenesis to drive the virus to extinction[123,129].

The conclusion of all these studies is that combination treatments should always be recommended for the management of chronic viral infections. Regarding mutagenic therapies, the same team concluded that a sequential approach, first using an inhibitor to reduce viral load, followed by treatment with a mutagen, can be more effective than administering both simultaneously

in driving the virus towards extinction[179]. The rationale is that in a population with a low viral load, it is more difficult for resistance mutations to arise and for the virus to expand and escaper the effect of mutagenic agents.

In our studies of chronically HEV-infected patients, a major limitation is that no alternative to RBV treatment exists, and combination therapies are still not available. From the two studies included in this thesis, we observed that in one patient, initial treatment with RBV was effective, and serum HEV-RNA became undetectable after several weeks of treatment. However, undetectable serum RNA does not necessary indicate complete viral clearance, as low-level replication may persist in the liver or other extra-hepatic sites. If RBV treatment is stopped immediately upon achieving upon achieving negative serum HEV-RNA, without maintaining RBV therapy, basal replication of HEV in the presence of declining RBV effective dose, can lead to viral relapse with the selection of a population more resistant to RBV. Based on our observations, we recommended that in chronically HEV-infected patients, RBV treatment should be continued for several months after serum RNA becomes negative, confirmed by at least to highly sensitive tests in both serum and feces. This strategy helps prevent the emergence of fitness-enhancing mutations and reduce the risk of treatment failure.

Furthermore, even in scenarios where a sequential treatment strategy is feasible, consisting in an initial antiviral inhibitor to reduce viral load followed by a mutagenic agent to induce lethal mutagenesis, the mutagenic treatment should be maintained for several months after serum RNA becomes undetectable using the most sensitive diagnostic assays. This prolonged treatment window is likely critical to prevent viral rebound and the re-expansion of fit variants capable of escaping mutagenic pressure.

6.3 Mechanisms Underlying Ribavirin Resistance in HEV Infection

Taken all, three possible solutions may account for the treatment failure observed across the three different RBV regimes, particularly for the patient described in the first study:

1.- As previously reported in a substantial proportion of patients chronically infected with hepatitis C virus (HCV) who failed direct-acting antiviral (DAA) therapy, treatment failure can occur in the absence of resistance-associated substitutions at appreciable frequencies. This phenomenon has been especially reported in patients with advanced liver damage, reflecting long-term Infection. We therefore considered whether a similar mechanism could underlie treatment failure in our two chronically HEV-infected patients.

2.- RBV-resistance substitutions may be selected during antiviral treatment and be responsible for treatment failure.

3.- A combination of both mechanisms. RBV treatment failure cannot be fully explained solely by the emergence of classical resistance mutations. Instead, multiple substitutions, not directly annotated as resistance-associated, may exert a compensatory effect that alters the configuration and geometry of the targeted protein in the vicinity of the drug-binding site, thereby reducing drug affinity and contributing to treatment failure.

To further characterize the emergence or presence of RAS under RBV treatment along the whole HEV genome, we conducted a third study (manuscript under revision) using the samples from the first study but applying a different sequencing and analysis methodology. Notably, due to the high variability of HEV, particularly under RBV treatment, our previous attempts to sequence the full HEV genome using an amplicon-based approach had failed.

Shotgun metagenomics using NGS was here employed to sequence the whole HEV genome. In this approach, in which the genome is fragmented into short sequences, QFF analysis is not feasible since different size fragments are generated. However, the rationale was to detect and compare substitutions across sequential samples collected during the different steps of the three RBV treatment regimen to determine whether resistance mutations were being selected. By applying shotgun metagenomics, we were able to recover the complete genome from samples with viral loads above 10^4 copies/mL. Full genome was successfully obtained from samples S1 (five days after infection confirmation), S2 (at initiation of the first treatment), S6 (after the end of the first treatment), S7 (at initiation of the second treatment)), and finally S17 (the final time point after following completion of all three RBV treatments).

The increase in variability during the infection course and successive RBV treatments was also shown with an increase in the Ts/Tv ratios and the number of total variants identified. What is more, two thirds of the variants that were kept in more than three study points (with AF > 0.05) appeared in S6 after the first treatment. Hence, this might reinforce the thesis of the mutagenic effect of RBV that was evidenced using the combination QFF and HNP.

However, despite this variability shown by the number of nucleotide variants (S1 to S17: 33, 30, 321, 173 and 603), we only detected 11 substitutions fixed at the phenotypic level. What is more, the fixed substitutions were not present in all samples, although most of them were present at AF > 0.05. This suggested that the variability is mostly presented within a sub-consensus level.

Among the RBV resistance–associated mutations previously reported in the literature, only Y1587F was detected at the consensus level in the final sampling time point (S17). In addition, a variant affecting the same residue as

the previously described D1384G mutation, namely D1384N, emerged at the consensus level in S17, resulting in the loss of the negatively charged amino acid. Notably, minor variant analysis revealed an alternative substitution at this same position (D1384T), highlighting residue 1384 as a potential hotspot for RBV resistance–associated variation. At a comparable allele frequency, the K1383N substitution was also detected. Importantly, the G1634R mutation was detected at time point S7 and persisted through S17.

To assess whether additional resistance-associated substitutions were present at low haplotype frequencies, variants with an allele frequency below 0.05 were examined. This analysis revealed that D1384N was already present at S7, well before its fixation at the consensus level, whereas K1398R emerged only at S17 after three courses of RBV therapy. Collectively, these findings indicate that several RBV resistance-associated mutations were detectable long before the end of treatment and prior to clinical failure. An in-depth NGS-based analysis of minor variants could therefore have anticipated the emergence and fixation of key resistance substitutions, such as D1384N and G1634R, potentially allowing early discontinuation of ineffective RBV therapy and more informed therapeutic decision-making.

Viruses are thought to have emerged early in the evolution of life and display an extraordinary capacity to parasitize a wide range of organisms across almost all taxonomic groups, spanning every domain of life. Their long-term evolutionary success relies on the ability to exploit a broad set of biological mechanisms available within their hosts. Given the intrinsic redundancy and adaptability of biological systems, it is plausible that viruses, and specially HEV, may employ multiple, non-exclusive strategies to ensure replication, persistence and transmission. In this sense, our results support that RBV-treatment failure can be explained by both the emergence of classical resistance mutations and an increase in viral diversity which is maintained even after stopping RBV treatment, probably incorporating compensatory mutations. These mutations enhance the fitness of the viruses carrying the

RAS by altering the configuration and geometry of the targeted protein near the drug-binding site, thereby reducing drug affinity and contributing to treatment failure.

7. CONCLUSIONS

1. The combination of the Quasispecies Fitness Fraction Partition (QFF) and the Hill Number Profile (HNP) provides an easily interpretable visualization of quasispecies evolution, which may have significant clinical relevance.
2. Quasispecies analysis using the novel QFF approach could serve as a useful tool to monitor viral populations ranging from very low to extremely high variability under mutagenic treatments during chronic disease progression.
3. Most RBV-induced genetic diversity consists of synonymous mutations, as the master phenotype remains predominantly preserved across all longitudinal samples.
4. There is a profound difference in how quasispecies responds to RBV at the genotypic versus phenotypic level. At the nucleotide level, the quasispecies become highly unstructured, showing a drastic reduction in the master sequence and a rise in rare haplotypes. However, at the amino acid level, the master phenotype remains highly prevalent and conserved, suggesting the virus maintains its functional fitness.
5. Ribavirin treatment, particularly at high viral loads, induces the selection of a diverse set of synonymous haplotypes. These synonymous variants act as reservoir of functional genomes that could replace the master sequence.
6. One clinical application of our results is that, during chronic HEV infection, initial RBV treatment in monotherapy should be continued even after serum RNA becomes undetectable for several weeks to prevent relapse. Otherwise, any relapse may promote the selection of fitness-enhancing mutations, ultimately leading to treatment failure.
7. Another clinical implication is that relapse following ineffective ribavirin treatment, which fails to achieve viral extinction, can increase viral diversity and carry the risk of selecting fitness-enhancing mutations. Therefore, mutagenic therapy may be most effective when combined with other antiviral inhibitors to minimize viral replication.

8. HEV exhibits significant genomic plasticity under sequential RBV pressure, with viral diversity and transition to transversion (Ts/Tv) ratio increasing over time.

9. Resistance variants to RBV could be detectable as minor variants at sub-consensus level at the moment of giving the second RBV treatment after the first viral rebound, long before they became dominant in the consensus sequence.

10. HEV might be able to adapt to ribavirin by exploiting the variability introduced and the expansion of the mutant cloud generated. The larger the mutant pool, the greater the probability of a resistance mutation emerging.

11. Relying solely on consensus sequence analysis is insufficient. Deep variant profiling provides significant clinical value for anticipating RBV treatment failure. Precision medicine approaches could enable clinicians to make more informed decisions and avoid ineffective treatments.

8. BIBLIOGRAPHY

1. Lemon, S. M. & Walker, C. M. Enterically Transmitted Non-A, Non-B Hepatitis and the Discovery of Hepatitis E Virus. *Cold Spring Harb Perspect Med* 9, a033449 (2019).

2. Blumberg, B. S., Alter, H. J. & Visnich, S. A New Antigen in Leukemia Sera. *JAMA* 191, 541–546 (1965).

3. Prince, A. M. An antigen detected in the blood during the incubation period of serum hepatitis. *Proc Natl Acad Sci U S A* 60, 814–821 (1968).

4. Feinstone, S. M., Kapikian, A. Z. & Purceli, R. H. Hepatitis A: Detection by immune electron microscopy of a viruslike antigen associated with acute illness. *Science (1979)* 182, 1026–1028 (1973).

5. Khuroo, M. S. Study of an epidemic of non-A, non-B hepatitis. *Am J Med* 68, 818–824 (1980).

6. Wong, D. C., Purcell, R. H., Mandyam Ammanjee Sreenivasan, Rama Prasad, S. & Pavri, K. M. Epidemic and endemic hepatitis in India: evidence for a non-A, non-B hepatitis virus aetiology. *The Lancet* 316, 876–879 (1980).

7. Khuroo, M. S. & Sofi, A. A. The Discovery of Hepatitis Viruses: Agents and Disease. *J Clin Exp Hepatol* 10, 391–401 (2020).

8. Balayan, M. S. *et al.* Evidence for a Virus in Non-A, Non-B Hepatitis Transmitted via the Fecal-Oral Route. *Intervirology* 20, 23–31 (1983).

9. Tam, A. W. *et al.* Hepatitis E virus (HEV): Molecular cloning and sequencing of the full-length viral genome. *Virology* 185, 120–131 (1991).

10. Bode Indu Reddy, and Surakaram, A. & M, and G. Hepatitis E Virus (HEV) Unveiled: A Comprehensive Examination from Genotypes, Epidemiology and Zoonotic Challenges to Public Health Strategies. in *Emerging Zoonotic Threats from Swine: A Public Health Perspective*

(eds. Deb Rajib, and Nayak, J., Singh, and S. G. & Kumar, and G. V.) 175–190 (Springer Nature Singapore, Singapore, 2025). doi:10.1007/978-981-96-7407-7_10.

11. Plümers, R. *et al.* Kinetics of Hepatitis E Virus Infections in Asymptomatic Persons - Volume 30, Number 5—May 2024 - Emerging Infectious Diseases journal - CDC. *Emerg Infect Dis* 30, 934–940 (2024).

12. Webb, G. W. & Dalton, H. R. Hepatitis E: an underestimated emerging threat. *Ther Adv Infect Dis* 6, 2049936119837162 (2019).

13. El-Mokhtar, M. A. *et al.* The Unmet Needs of Hepatitis E Virus Diagnosis in Suspected Drug-Induced Liver Injury in Limited Resource Setting. *Front Microbiol* 12, 737486 (2021).

14. Sauleda, S. *et al.* Seroprevalence of hepatitis E virus (HEV) and detection of HEV RNA with a transcription-mediated amplification assay in blood donors from Catalonia (Spain). *Transfusion (Paris)* 55, 972–979 (2015).

15. Sridhar, S., Teng, J. L. L., Chiu, T. H., Lau, S. K. P. & Woo, P. C. Y. Hepatitis E Virus Genotypes and Evolution: Emergence of Camel Hepatitis E Variants. *Int J Mol Sci* 18, (2017).

16. Purdy, M. A. *et al.* ICTV Virus Taxonomy Profile: Hepeviridae. *J Gen Virol* 98, 2645–2646 (2017).

17. Kamar, N. *et al.* Hepatitis E virus infection. *Nat Rev Dis Primers* 3, 17086 (2017).

18. Boxman, I. L. A. *et al.* Monitoring of pork liver and meat products on the Dutch market for the presence of HEV RNA. *Int J Food Microbiol* 296, 58–64 (2019).

19. Acosta, J. *et al.* Zoonotic transmission of hepatitis E virus in a pig farmer from Argentina: A case report. *Zoonoses Public Health* https://doi.org/10.1111/zph.12902 (2021) doi:10.1111/zph.12902.

20. Intharasongkroh, D. *et al.* Hepatitis E Virus in Pork and Variety Meats Sold in Fresh Markets. *Food Environ Virol* 9, 45–53 (2017).

21. Riveiro-Barciela, M. *et al.* Phylogenetic demonstration of hepatitis E infection transmitted by pork meat ingestion. *J Clin Gastroenterol* 49, 165–168 (2015).

22. Hoofnagle, J. H., Nelson, K. E. & Purcell, R. H. Hepatitis E. *N Engl J Med* 367, 1237–1244 (2012).

23. Riveiro-Barciela, M. *et al.* Thrombotic thrombocytopenic purpura relapse induced by acute hepatitis E transmitted by cryosupernatant plasma and successfully controlled with ribavirin. *Transfusion (Paris)* 58, 2501–2505 (2018).

24. Bes, M. *et al.* Effect of Hepatitis E Virus RNA Universal Blood Donor Screening, Catalonia, Spain, 2017–2020. *Emerg Infect Dis* 28, 157–165 (2022).

25. Riveiro-Barciela, M. *et al.* Red blood cell transfusion-transmitted acute hepatitis E in an immunocompetent subject in Europe: a case report. *Transfusion (Paris)* 57, 244–247 (2017).

26. Takahashi, M. *et al.* Molecular characterization of a novel hepatitis E virus (HEV) strain obtained from a wild boar in Japan that is highly divergent from the previously recognized HEV strains. *Virus Res* 180, 59–69 (2014).

27. Woo, P. C. Y. *et al.* New Hepatitis E Virus Genotype in Camels, the Middle East. *Emerg Infect Dis* 20, 1044 (2014).

28. Lee, G. H. *et al.* Chronic Infection with Camelid Hepatitis e Virus in a Liver Transplant Recipient Who Regularly Consumes Camel Meat and Milk. *Gastroenterology* 150, 355-357.e3 (2016).

29. Woo, P. C. Y. *et al.* New Hepatitis E Virus Genotype in Bactrian Camels, Xinjiang, China, 2013. *Emerg Infect Dis* 22, 2219 (2016).

30. Purdy, M. A. *et al.* ICTV Virus Taxonomy Profile: Hepeviridae. *J Gen Virol* 98, 2645–2646 (2017).

31. Sun, P., Lin, S., He, S., Zhou, E. M. & Zhao, Q. Avian hepatitis E virus: With the trend of genotypes and host expansion. *Front Microbiol* 10, 470999 (2019).

32. Dinkelborg, K., Lhomme, S. & Gömer, A. Rat hepatitis E virus: An emerging challenge in human hepatitis. *J Hepatol* 84, 10–12 (2025).

33. Zahmanova, G. *et al.* The Re-Emergence of Hepatitis E Virus in Europe and Vaccine Development. *Viruses 2023, Vol. 15, Page 1558* 15, 1558 (2023).

34. Feng, Z., Hirai-Yuki, A., McKnight, K. L. & Lemon, S. M. Naked Viruses That Aren't Always Naked: Quasi-Enveloped Agents of Acute Hepatitis. *Annu Rev Virol* 1, 539–560 (2014).

35. Nagashima, S. *et al.* Characterization of the Quasi-Enveloped Hepatitis E Virus Particles Released by the Cellular Exosomal Pathway. *J Virol* 91, (2017).

36. Takahashi, M. *et al.* Hepatitis E Virus (HEV) Strains in Serum Samples Can Replicate Efficiently in Cultured Cells Despite the Coexistence of HEV Antibodies: Characterization of HEV Virions in Blood Circulation. *J Clin Microbiol* 48, 1112 (2010).

37. Allweiss, L. *et al.* Human liver chimeric mice as a new model of chronic hepatitis e virus infection and preclinical drug evaluation. *J Hepatol* 64, 1033–1040 (2016).

38. Costafreda, M. I., Sauleda, S., Rico, A., Piron, M. & Bes, M. Detection of Nonenveloped Hepatitis E Virus in Plasma of Infected Blood Donors. *J Infect Dis* 226, 1753–1760 (2022).

39. Nair, V. P. *et al.* Endoplasmic Reticulum Stress Induced Synthesis of a Novel Viral Factor Mediates Efficient Replication of Genotype-1 Hepatitis E Virus. *PLoS Pathog* 12, e1005521 (2016).

40. Shahini, E. *et al.* Hepatitis E Virus: What More Do We Need to Know? *Medicina 2024, Vol. 60, Page 998* 60, 998 (2024).

41. Magden, J. *et al.* Virus-specific mRNA capping enzyme encoded by hepatitis E virus. *J Virol* 75, 6249–6255 (2001).

42. Parvez, M. K. Mutational analysis of hepatitis E virus ORF1 'Y-domain': Effects on RNA replication and virion infectivity. *World J Gastroenterol* 23, 590 (2017).

43. Parvez, M. K. & Khan, A. A. Molecular modeling and analysis of hepatitis E virus (HEV) papain-like cysteine protease. *Virus Res* 179, 220 (2013).

44. Paliwal, D., Panda, S. K., Kapur, N., Varma, S. P. K. & Durgapal, H. Hepatitis E virus (HEV) protease: A chymotrypsin-like enzyme that processes both non-structural (pORF1) and capsid (pORF2) protein. *Journal of General Virology* 95, 1689–1700 (2014).

45. Glitscher, M. *et al.* The Protease Domain in HEV pORF1 Mediates the Replicase's Localization to Multivesicular Bodies and Its Exosomal Release. *Cell Mol Gastroenterol Hepatol* 17, 589–605 (2024).

46. Wißing, M. H. *et al.* Genetic determinants of host- and virus-derived insertions for hepatitis E virus replication. *Nat Commun* 15, 4855 (2024).

47. Martien, M. R. *et al.* A Biochemical Genomics Approach for Identifying Genes by the Activity of Their Products. *Science (1979)* 286, 1153–1155 (1999).

48. Lhomme, S. *et al.* Influence of Polyproline Region and Macro Domain Genetic Heterogeneity on HEV Persistence in Immunocompromised Patients. *J Infect Dis* 209, 300 (2013).

49. Anang, S. *et al.* Identification of critical residues in Hepatitis e virus macro domain involved in its interaction with viral methyltransferase and ORF3 proteins. *Sci Rep* 6, 25133- (2016).

50. Ju Xiaohui and Dong, L. and D. Q. Hepatitis E Virus Life Cycle. in *Hepatitis E Virus* (ed. Wang, Y.) 141–157 (Springer Nature Singapore, Singapore, 2023). doi:10.1007/978-981-99-1304-6_10.

51. Guu, T. S. Y. *et al.* Structure of the hepatitis E virus-like particle suggests mechanisms for virus assembly and receptor binding. *Proc Natl Acad Sci U S A* 106, 12992 (2009).

52. Yamashita, T. *et al.* Biological and immunological characteristics of hepatitis E virus-like particles based on the crystal structure. *Proc Natl Acad Sci U S A* 106, 12986–12991 (2009).

53. Yin, X. *et al.* Origin, antigenicity, and function of a secreted form of ORF2 in hepatitis E virus infection. *Proc Natl Acad Sci U S A* 115, 4773–4778 (2018).

54. Graff, J. *et al.* Mutations within potential glycosylation sites in the capsid protein of hepatitis E virus prevent the formation of infectious virus particles. *J Virol* 82, 1185–1194 (2008).

55. Shafat, Z., Islam, A. & Parveen, S. Amino acid pattern reveals multi-functionality of ORF3 protein from HEV. *Bioinformation* 20, 121 (2024).

56. Nagashima, S. *et al.* The membrane on the surface of hepatitis E virus particles is derived from the intracellular membrane and contains trans-Golgi network protein 2. *Arch Virol* 159, 979–991 (2014).

57. Nan, Y. & Zhang, Y. J. Molecular biology and infection of hepatitis E virus. *Front Microbiol* 7, 215637 (2016).

58. Okamoto, H. Culture systems for hepatitis E virus. *Journal of Gastroenterology 2012 48:2* 48, 147–158 (2012).

59. Das, A. *et al.* Cell entry and release of quasi-enveloped human hepatitis viruses. *Nat Rev Microbiol* 21, 573–589 (2023).

60. Marion, O. *et al.* Hepatitis E virus replication in human intestinal cells. *Gut* 69, 901–910 (2020).

61. Capelli, N. *et al.* Vectorial Release of Hepatitis E Virus in Polarized Human Hepatocytes. *J Virol* 93, (2019).

62. Yin, X., Ambardekar, C., Lu, Y. & Feng, Z. Distinct Entry Mechanisms for Nonenveloped and Quasi-Enveloped Hepatitis E Viruses. *J Virol* 90, 4232–4242 (2016).

63. Kapur, N., Thakral, D., Durgapal, H. & Panda, S. K. Hepatitis e virus enters liver cells through receptor-dependent clathrin-mediated endocytosis. *J Viral Hepat* 19, 436–448 (2012).

64. Holla, P., Ahmad, I., Ahmed, Z. & Jameel, S. Hepatitis E virus enters liver cells through a dynamin-2, clathrin and membrane cholesterol-dependent pathway. *Traffic* 16, 398–416 (2015).

65. Yin, X., Li, X. & Feng, Z. Role of Envelopment in the HEV Life Cycle. *Viruses* 8, 229 (2016).

66. Kalia, M., Chandra, V., Rahman, S. A., Sehgal, D. & Jameel, S. Heparan sulfate proteoglycans are required for cellular binding of the hepatitis E virus ORF2 capsid protein and for viral infection. *J Virol* 83, 12714–12724 (2009).

67. Zhang, L. *et al.* Asialoglycoprotein receptor facilitates infection of PLC/PRF/5 cells by HEV through interaction with ORF2. *J Med Virol* 88, 2186–2195 (2016).

68. Yu, H. *et al.* Homology model and potential virus-capsid binding site of a putative HEV receptor Grp78. *J Mol Model* 17, 987–995 (2011).

69. Jemielity, S. *et al.* TIM-family proteins promote infection of multiple enveloped viruses through virion-associated phosphatidylserine. *PLoS Pathog* 9, e1003232 (2013).

70. Suppiah, S., Zhou, Y. & Frey, T. K. Lack of processing of the expressed ORF1 gene product of hepatitis E virus. *Virol J* 8, 245 (2011).

71. Sehgal, D., Thomas, S., Chakraborty, M. & Jameel, S. Expression and processing of the Hepatitis E virus ORF1 nonstructural polyprotein. *Virology Journal 2006 3:1* 3, 38- (2006).

72. Cancela, F., Noceti, O., Arbiza, J. & Mirazo, S. Structural aspects of hepatitis E virus. *Arch Virol* 167, 2457–2481 (2022).

73. Hervouet, K. *et al.* An Arginine-Rich Motif in the ORF2 capsid protein regulates the hepatitis E virus lifecycle and interactions with the host cell. *PLoS Pathog* 18, e1010798 (2022).

74. Montpellier, C. *et al.* Hepatitis E Virus Lifecycle and Identification of 3 Forms of the ORF2 Capsid Protein. *Gastroenterology* 154, 211-223.e8 (2018).

75. Feng, Z. & Lemon, S. M. Peek-a-boo: Membrane hijacking and the pathogenesis of viral hepatitis. *Trends Microbiol* 22, 59–64 (2014).

76. Nagashima, S. *et al.* Tumour susceptibility gene 101 and the vacuolar protein sorting pathway are required for the release of hepatitis E virions. *Journal of General Virology* 92, 2838–2848 (2011).

77. Nagashima, S. *et al.* A PSAP motif in the ORF3 protein of hepatitis E virus is necessary for virion release from infected cells. *Journal of General Virology* 92, 269–278 (2011).

78. Gouttenoire, J. *et al.* Palmitoylation mediates membrane association of hepatitis E virus ORF3 protein and is required for infectious particle secretion. *PLoS Pathog* 14, e1007471 (2018).

79. Tian, D. *et al.* Hepatitis E virus infects brain microvascular endothelial cells, crosses the blood–brain barrier, and invades the central nervous system. *Proc Natl Acad Sci U S A* 119, e2201862119 (2022).

80. Hoofnagle, J. H., Nelson, K. E. & Purcell, R. H. Hepatitis E. *New England Journal of Medicine* 367, 1237–1244 (2012).

81. Aggarwal, R., Kini, D., Sofat, S., Naik, S. R. & Krawczynski, K. Duration of viraemia and faecal viral excretion in acute hepatitis E. *Lancet* 356, 1081–1082 (2000).

82. Wedemeyer, H., Pischke, S. & Manns, M. P. Pathogenesis and Treatment of Hepatitis E Virus Infection. *Gastroenterology* 142, 1388-1397.e1 (2012).

83. Dalton, H. R., Bendall, R., Ijaz, S. & Banks, M. Hepatitis E: an emerging infection in developed countries. *Lancet Infect Dis* 8, 698–709 (2008).

84. Blasco-Perrin, H. *et al.* Hepatitis E virus in patients with decompensated chronic liver disease: a prospective UK/French study. *Aliment Pharmacol Ther* 42, 574–581 (2015).

85. Kumar, A. & Saraswat, V. A. Hepatitis E and Acute-on-Chronic Liver Failure. *J Clin Exp Hepatol* 3, 225–230 (2013).

86. Shalimar & Acharya, S. K. Hepatitis E and Acute Liver Failure in Pregnancy. *J Clin Exp Hepatol* 3, 213–224 (2013).

87. Jilani, N. *et al.* Hepatitis E virus infection and fulminant hepatic failure during pregnancy. *J Gastroenterol Hepatol* 22, 676–682 (2007).

88. Zhu, F. C. *et al.* Efficacy and safety of a recombinant hepatitis e vaccine in healthy adults: A large-scale, randomised, double-blind placebo-controlled, phase 3 trial. *The Lancet* 376, 895–902 (2010).

89. Guillois, Y. *et al.* High Proportion of Asymptomatic Infections in an Outbreak of Hepatitis E Associated With a Spit-Roasted Piglet, France, 2013. *Clin Infect Dis* 62, 351–357 (2016).

90. Faber, M. S. *et al.* Hepatitis E virus seroprevalence among adults, Germany. *Emerg Infect Dis* 18, 1654–1657 (2012).

91. Aherfi, S. *et al.* Liver transplantation for acute liver failure related to autochthonous genotype 3 hepatitis E virus infection. *Clin Res Hepatol Gastroenterol* 38, 24–31 (2014).

92. Manka, P. *et al.* Hepatitis E Virus Infection as a Possible Cause of Acute Liver Failure in Europe. *Clinical Gastroenterology and Hepatology* 13, 1836-1842.e2 (2015).

93. Anastasiou, O. E., Thodou, V., Berger, A., Wedemeyer, H. & Ciesek, S. Comprehensive Evaluation of Hepatitis E Serology and Molecular Testing in a Large Cohort. *Pathogens* 9, 137 (2020).

94. Abravanel, F. *et al.* Hepatitis E virus reinfections in solid-organ-transplant recipients can evolve into chronic infections. *J Infect Dis* 209, 1900–1906 (2014).

95. Zhang, J. *et al.* Protection against hepatitis E virus infection by naturally acquired and vaccine-induced immunity. *Clinical Microbiology and Infection* 20, O397–O405 (2014).

96. Aggarwal, R. Diagnosis of hepatitis E. *Nature Reviews Gastroenterology & Hepatology 2012 10:1* 10, 24–33 (2012).

97. Baylis, S. A., Hanschmann, K. M., Blümel, J. & Nübling, C. M. Standardization of hepatitis E virus (HEV) nucleic acid amplification technique-based assays: an initial study to evaluate a panel of HEV strains and investigate laboratory performance. *J Clin Microbiol* 49, 1234–1239 (2011).

98. Kamar, N., Rostaing, L., Legrand-Abravanel, F. & Izopet, J. How should hepatitis e virus infection be defined in organ-transplant recipients? *American Journal of Transplantation* 13, 1935–1936 (2013).

99. Kamar, N. *et al.* Hepatitis E Virus and Chronic Hepatitis in Organ-Transplant Recipients. *New England Journal of Medicine* 358, 811–817 (2008).

100. Mikulska, M. *et al.* HEV infection in stem cell transplant recipients—retrospective study of EBMT Infectious Diseases Working Party. *Bone Marrow Transplantation 2021 57:2* 57, 167–175 (2021).

101. Oilier, L. *et al.* Chronic hepatitis after hepatitis E virus infection in a patient with non-Hodgkin lymphoma taking rituximab. *Ann Intern Med* 150, 430–431 (2009).

102. Dalton, H. R., Bendall, R. P., Keane, F. E., Tedder, R. S. & Ijaz, S. Persistent carriage of hepatitis E virus in patients with HIV infection. *N Engl J Med* 361, 1025–1027 (2009).

103. Geng, Y. *et al.* Persistent hepatitis e virus genotype 4 infection in a child with acute lymphoblastic leukemia. *Hepat Mon* 14, e15618 (2014).

104. Kamar, N. *et al.* Hepatitis E virus-related cirrhosis in kidney-and kidney-pancreas- transplant recipients. *American Journal of Transplantation* 8, 1744–1748 (2008).

105. Kamar, N. *et al.* Influence of immunosuppressive therapy on the natural history of genotype 3 hepatitis-E virus infection after organ transplantation. *Transplantation* 89, 353–360 (2010).

106. Kamar, N. *et al.* Factors associated with chronic hepatitis in patients with hepatitis E virus infection who have received solid organ transplants. *Gastroenterology* 140, 1481–1489 (2011).

107. Suneetha, P. V. *et al.* Hepatitis E virus (HEV)-specific T-cell responses are associated with control of HEV infection. *Hepatology* 55, 695–708 (2012).

108. Wu, J. *et al.* Extrahepatic manifestations related to hepatitis E virus infection and their triggering mechanisms. *Journal of Infection* 83, 298–305 (2021).

109. Fousekis, F. S., Mitselos, I. V. & Christodoulou, D. K. Extrahepatic manifestations of hepatitis E virus: An overview. *Clin Mol Hepatol* 26, 16 (2019).

110. Fukae, J. *et al.* Guillain-Barré and Miller Fisher syndromes in patients with anti-hepatitis E virus antibody: a hospital-based survey in Japan. *Neurol Sci* 37, 1849–1851 (2016).

111. Van Den Berg, B. *et al.* Guillain-Barré syndrome associated with preceding hepatitis E virus infection. *Neurology* 82, 491–497 (2014).

112. Van Eijk, J. J. J. *et al.* Neuralgic amyotrophy and hepatitis E virus infection. *Neurology* 82, 498–503 (2014).

113. Silva, M. *et al.* Hepatitis E virus infection as a direct cause of neuralgic amyotrophy. *Muscle Nerve* 54, 325–327 (2016).

114. Murkey, J. A. *et al.* Hepatitis E Virus-Associated Meningoencephalitis in a Lung Transplant Recipient Diagnosed by Clinical Metagenomic Sequencing. *Open Forum Infect Dis* 4, ofx121 (2017).

115. Guinault, D. *et al.* Hepatitis e virus-induced cryoglobulinemic glomerulonephritis in a nonimmunocompromised person. *American Journal of Kidney Diseases* 67, 660–663 (2016).

116. Singh, N. K. & Gangappa, M. Acute immune thrombocytopenia associated with hepatitis E in an adult. *Am J Hematol* 82, 942–943 (2007).

117. Rauff, B. *et al.* Hepatitis associated aplastic anemia: a review. *Virol J* 8, 87 (2011).

118. Deng, H., Xu, M., Liu, H. & Gao, H. Clinical analysis of the characteristics of patients diagnosed with hemolytic anemia induced by hepatitis E viral infection. *iLIVER* 1, 117–121 (2022).

119. Dietz, J. *et al.* Deep sequencing reveals mutagenic effects of ribavirin during monotherapy of hepatitis C virus genotype 1-infected patients. *J Virol* 87, 6172–6181 (2013).

120. Tejada, S. *et al.* Ribavirin for Treatment of Subjects with Respiratory Syncytial Virus-Related Infection: A Systematic Review and Meta-Analysis. *Adv Ther* 39, 4037–4051 (2022).

121. Salam, A. P. *et al.* Ribavirin for treating Lassa fever: A systematic review of pre-clinical studies and implications for human dosing. *PLoS Negl Trop Dis* 16, (2022).

122. Soares-Weiser, K., Thomas, S., G, G. T. & Garner, P. Ribavirin for Crimean-Congo hemorrhagic fever: systematic review and meta-analysis. *BMC Infect Dis* 10, (2010).

123. Domingo, E., Escarmis, C., Lazaro, E. & Manrubia, S. C. Quasispecies dynamics and RNA virus extinction. *Virus Res.* 107, 129–139 (2005).

124. Graci, J. D. & Cameron, C. E. Mechanisms of action of ribavirin against distinct viruses. *Rev Med Virol* 16, 37–48 (2006).

125. Dalton, H. R. *et al.* EASL Clinical Practice Guidelines on hepatitis E virus infection. *J Hepatol* 68, 1256–1271 (2018).

126. Rivero-Juarez, A. *et al.* Ribavirin as a first treatment approach for hepatitis e virus infection in transplant recipient patients. *Microorganisms* 8, 51 (2020).

127. Kamar, N. *et al.* Ribavirin for Chronic Hepatitis E Virus Infection in Transplant Recipients. *New England Journal of Medicine* 370, 1111–1120 (2014).

128. Debing, Y. *et al.* Ribavirin inhibits in vitro hepatitis E virus replication through depletion of cellular GTP pools and is moderately synergistic with alpha interferon. *Antimicrob Agents Chemother* 58, 267–273 (2014).

129. Perales, C. *et al. The Increasing Impact of Lethal Mutagenesis of Viruses*. *Future Medicinal Chemistry* vol. 11 1645–1657 (Future Medicine Ltd., 2019).

130. Hadj Hassine, I., Ben M'Hadheb, M. & Menéndez-Arias, L. Lethal Mutagenesis of RNA Viruses and Approved Drugs with Antiviral Mutagenic Activity. *Viruses* 14, 841 (2022).

131. Wahid, A. *et al.* Extrahepatic Replication and Genomic Signatures of the Hepatitis E Virus in the Kidney. *Liver Int* 45, (2025).

132. Todt, D. *et al.* In vivo evidence for ribavirin-induced mutagenesis of the hepatitis E virus genome. *Gut* 65, 1733–1743 (2016).

133. van Tong, H. *et al.* Hepatitis E Virus Mutations: Functional and Clinical Relevance. *EBioMedicine* 11, 31–42 (2016).

134. Debing, Y. *et al.* Hepatitis E virus mutations associated with ribavirin treatment failure result in altered viral fitness and ribavirin sensitivity. *J Hepatol* 65, 499–508 (2016).

135. Sastre, L. *et al.* The challenge of chronic hepatitis E in liver transplant recipients: Failure of sofosbuvir plus ribavirin therapy. *Gastroenterol Hepatol* 43, 136–137 (2020).

136. Schulz, M. *et al.* Combination therapy of sofosbuvir and ribavirin fails to clear chronic hepatitis E infection in a multivisceral transplanted patient. *Journal of Hepatology* vol. 71 225–227 Preprint at https://doi.org/10.1016/j.jhep.2019.03.029 (2019).

137. Lampejo, T. Sofosbuvir in the Treatment of Hepatitis E virus Infection: A Review of in vitro and in vivo Evidence. *J Clin Exp Hepatol* 12, 1225–1237 (2022).

138. Gömer, A. *et al.* Emergence of resistance-associated variants during sofosbuvir treatment in chronically infected hepatitis E patients. *Hepatology* 78, 1882–1895 (2023).

139. Gömer, A. *et al.* Dynamic evolution of the sofosbuvir-associated variant A1343V in HEV-infected patients under concomitant sofosbuvir-ribavirin treatment. *J Hepatol* 6, 100989 (2024).

140. Kamar, N., Del Bello, A., Abravanel, F., Pan, Q. & Izopet, J. Unmet Needs for the Treatment of Chronic Hepatitis E Virus Infection in Immunocompromised Patients. *Viruses* 14, 2116 (2022).

141. Netzler, N. E., Tuipulotu, D. E., Vasudevan, S. G., Mackenzie, J. M. & White, P. A. Antiviral Candidates for Treating Hepatitis E Virus Infection. *Antimicrob Agents Chemother* 63, (2019).

142. Jagst, M., Gömer, A., Todt, D. & Steinmann, E. Performance of sofosbuvir and NITD008 in extrahepatic neuronal cells against HEV. *Antiviral Res* 227, 105922 (2024).

143. Hooda, P., Al-Dosari, M., Sinha, N., Parvez, M. K. & Sehgal, D. Inhibition of HEV Replication by FDA-Approved RdRp Inhibitors. *ACS Omega* 8, 41570–41578 (2023).

144. Kaushik, N. *et al.* Zinc Salts Block Hepatitis E Virus Replication by Inhibiting the Activity of Viral RNA-Dependent RNA Polymerase. *J Virol* 91, 754–771 (2017).

145. Marion, O., Abravanel, F., Izopet, J. & Kamar, N. Failure to respond to ribavirin despite elevated intra-erythrocyte zinc level in transplant-patients with chronic hepatitis E virus infection. *Transplant Infectious Disease* 21, e13050 (2019).

146. Todt, D. *et al.* The natural compound silvestrol inhibits hepatitis E virus (HEV) replication in vitro and in vivo. *Antiviral Res* 157, 151–158 (2018).

147. Praditya, D. F. *et al.* Identification of structurally re-engineered rocaglates as inhibitors against hepatitis E virus replication. *Antiviral Res* 204, 105359 (2022).

148. Nimgaonkar, I. *et al.* Isocotoin suppresses hepatitis E virus replication through inhibition of heat shock protein 90. *Antiviral Res* 185, 104997 (2021).

149. Domingo, E. & Holland, J. J. Mutation rates and rapid evolution of RNA viruses. in *Evolutionary biology of viruses* (ed. Morse, S. S.) 161–184 (Raven Press, New York, 1994).

150. Domingo, E., García-Crespo, C., Lobo-Vega, R. & Perales, C. Mutation Rates, Mutation Frequencies, and Proofreading-Repair Activities in RNA Virus Genetics. *Viruses* 13, 1882 (2021).

151. Domingo, E. & Holland, J. J. High error rates, population equilibrium, and evolution of RNA replication systems. in *RNA Genetics* (eds. Domingo, E. & Holland, J. J.) 3–36 (CRC Press, Boca Raton, Florida, 1988).

152. Domingo, E., Sheldon, J. & Perales, C. Viral Quasispecies Evolution. *Microbiol Mol Biol Rev* 76, 159 (2012).

153. Eigen, M. & Schuster, P. The hypercycle. A principle of natural self-organization. Part A: Emergence of the hypercycle. *Naturwissenschaften* 64, 541–565 (1977).

154. Summers, J. & Litwin, S. Examining The Theory of Error Catastrophe. *J Virol* 80, 20 (2006).

155. Eigen, M. *et al.* Molecular quasi-species. *Journal of Physical Chemistry* 92, 6881–6891 (2002).

156. Swanstrom, R. & Schinazi, R. F. Lethal mutagenesis as an antiviral strategy. *Science (1979)* 375, 497–498 (2022).

157. Gregori, J. *et al.* Inference with viral quasispecies diversity indices: clonal and NGS approaches. *Bioinformatics* 30, 1104–1111 (2014).

158. Beerenwinkel, N., Günthard, H. F., Roth, V. & Metzner, K. J. Challenges and opportunities in estimating viral genetic diversity from next-generation sequencing data. *Front Microbiol* 3, 31515 (2012).

159. Perales, C. *et al.* Baseline hepatitis C virus resistance-associated substitutions present at frequencies lower than 15% may be clinically significant. *Infect Drug Resist* 11, 2207–2210 (2018).

160. Lu, I. N., Muller, C. P. & He, F. Q. Applying next-generation sequencing to unravel the mutational landscape in viral quasispecies. *Virus Res* 283, 197963 (2020).

161. Gregori, J. *et al.* Viral quasispecies complexity measures. *Virology* 493, 227–237 (2016).

162. Gotelli, N. J. & Chao, A. Measuring and Estimating Species Richness, Species Diversity, and Biotic Similarity from Sampling Data. *Encyclopedia of Biodiversity* 5, 195–211 (2013).

163. Nei, M. *Molecular Evolutionary Genetics*. (Columbia University Press, New York, 1987).

164. Hill, M. O. Diversity and Evenness: A Unifying Notation and Its Consequences. *Ecology* 54, 427–432 (1973).

165. Gregori, J. *et al.* Rare haplotype load as marker for lethal mutagenesis. *PLoS One* 13, e0204877 (2018).

166. Quer, J. *et al.* Deep sequencing in the management of hepatitis virus infections. *Virus Res* 239, 115–125 (2017).

167. Martell, M. *et al.* Hepatitis C virus (HCV) circulates as a population of different but closely related genomes: Quasispecies nature of HCV genome distribution. *J Virol* 66, 3225–3229 (1992).

168. Gregori, J. *et al.* Ultra-deep pyrosequencing (UDPS) data treatment to study amplicon HCV minor variants. *PLoS One* 8, e83361 (2013).

169. Chen, Q. *et al.* Deep-sequencing reveals broad subtype-specific HCV resistance mutations associated with treatment failure. *Antiviral Res* 174, 104694 (2020).

170. Rodriguez-Frias, F. *et al.* High HCV subtype heterogeneity in a chronically infected general population revealed by high-resolution hepatitis C virus subtyping. *Clinical Microbiology and Infection* 23, p775.e1-775.e6 (2017).

171. Garcia-Cehic, D. *et al.* Resistance-associated substitutions after sofosbuvir/velpatasvir/voxilaprevir triple therapy failure. *J Viral Hepat* 28, 1319–1324 (2021).

172. Takahashi, K. *et al.* Estimation of the mutation rate of hepatitis E virus based on a set of closely related 7.5-year-apart isolates from Sapporo, Japan. *Hepatology Research* 29, 212–215 (2004).

173. Hartard, C. *et al.* Variability in molecular characteristics of Hepatitis E virus quasispecies could modify viral surface properties and transmission. *J Viral Hepat* 28, 1078–1090 (2021).

174. Melchert, J. *et al.* Whole genome sequencing reveals insights into hepatitis E virus genome diversity, and virus compartmentalization in chronic hepatitis E. *J Clin Virol* 168, 105583 (2023).

175. Agarwal, S., Baccam, P., Aggarwal, R. & Veerapu, N. S. Novel Synthesis and Phenotypic Analysis of Mutant Clouds for Hepatitis E Virus Genotype 1. *J Virol* 92, e01932-17 (2018).

176. Grandadam, M. *et al.* Evidence for hepatitis E virus quasispecies. *Journal of General Virology* 85, 3189–3194 (2004).

177. Lhomme, S. *et al.* Hepatitis E Virus Quasispecies and the Outcome of Acute Hepatitis E in Solid-Organ Transplant Patients. *J Virol* 86, 10006 (2012).

178. Abravanel, F. *et al.* Hepatitis e virus quasispecies in cerebrospinal fluid with neurological manifestations. *Vaccines (Basel)* 9, 1205 (2021).

179. Perales, C., Agudo, R., Tejero, H., Manrubia, S. C. & Domingo, E. Potential benefits of sequential inhibitor-mutagen treatments of RNA virus infections. *PLoS Pathog.* 5, e1000658 (2009).