INSTITUTO DE HIGIENE E MEDICINA TROPICAL
DESDE 1902

Universidade Nova de Lisboa

Instituto de Higiene e Medicina Tropical

Evolutionary History and Phylogeography of the hepatitis C and hepatitis B viruses in Portugal

Rute Alexandra Carvalho Antunes Marcelino

Dissertation presented to fulfill the requirements necessary to obtain a Ph.D. degree in Human Genetics and Infectious Diseases

(JANUARY, 2022)

Universidade Nova de Lisboa

Instituto de Higiene e Medicina Tropical

Evolutionary History and Phylogeography of the hepatitis C and hepatitis B viruses in Portugal

Author: Rute Alexandra Carvalho Antunes Marcelino

Supervisor: Doctor José Marcelino

Co-supervisors: Professor Dr. Ana Barroso Abecasis and Professor Dr. Rui Tato Marinho

Dissertation presented to fulfill the requirements necessary to obtain a Ph.D. degree in Human Genetics and Infectious Diseases

SCIENTIC PRODUCTION ASSOCIATED TO THIS STUDY

Papers published in peer reviewed international journals:

Rubén Brandão, **Rute Marcelino**, Fátima Gonçalves, Isabel Diogo, Ana Carvalho, Joaquim Cabanas, Inês Costa, Pedro Brogueira, Fernando Ventura, Ana Miranda, Kamal Mansinho and Perpétua Gomes. **Characterization of *NS5A* and *NS5B* Resistance-Associated Substitutions from Genotype 1 Hepatitis C Virus Infected Patients in a Portuguese Cohort.** Viruses. 2018 Apr 26;10(5):223. doi: 10.3390/v10050223.
Impact Factor: 3.816

Claudia Palladino, Ifeanyi Jude Ezeonwumelu, **Rute Marcelino**, Veronica Briz, Inês Moranguinho, Fátima Serejo, José Fernando Velosa, Rui Tato Marinho, Pedro Borrego and Nuno Taveira. **Epidemic history of hepatitis C virus genotypes in Portugal**. Scientific Reports. 2018 Aug 16;8(19): 12266. doi: 10.1038/s41598-018-30528-0
Impact factor: 3.998

Papers submitted as first author:

Rute Marcelino, Ifeanyi Jude Ezeonwumelu, André Janeiro, Paula Mimoso, Sónia Matos, Veronica Briz, Rui Tato Marinho, José Maria Marcelino, Nuno Taveira and Ana Abecasis. **Phylogeography of hepatitis B: The role of Portugal in the dissemination of HBV worldwide**. Submitted to MedRxiv 2022 jan 05**. (Preprint 10.1101/2022.01.05.22268725).**

Completed papers ready for submission to peer reviewed international journals:

Rute Marcelino, Ifeanyi Jude Ezeonwumelu, Rubén Brandão, Veronica Briz, Pedro Brogueira, Fernando Ventura, Ana Miranda, Kamal Mansinho, Perpétua Gomes, Rui Tato Marinho, José Maria Marcelino, Nuno Taveira and Ana Abecasis. **Reconstruction of the Spatial and Temporal Dynamics of Clade I and II of Hepatitis C Virus Subtype 1a in Portugal**. To submit to Viruses, 2022 Feb

Impact Factor: 5.048

Oral Communications in national conferences:

IX Jornadas do Instituto de Higiene e Medicina Tropical; 12th December 2018; "Epidemiological History and Susceptibility of HCV to Antivirals in Portugal".

Poster Presentation:

16th European AIDS conference; 25th-27th October, 2017; Milan, Italy; Characterization of NS5 coding region RASs from DAA-naïve GT1 HCV infected patients in Portugal.

Semana Digestiva; 20th-23rd June, 2018; Alfândega do Porto, Portugal; Diversity and NS5B natural polymorphisms in HCV viruses driving the current epidemic in Portugal.

Encontro Ciência 2018; 2nd-4th July, 2018; Centro de Congressos de Lisboa, Portugal; Epidemiological History and Susceptibility to hepatitis C virus antivirals in Portugal.

7th International Symposium on Hepatitis Care on Substance Users; 18th-21st September, 2018. Lisbon, Portugal. Origin, Diversity and Transmission Dynamics of Hepatitis C Virus Genotypes in Portugal.

X Jornadas do Instituto de Higiene e Medicina Tropical; 12th December, 2019; Lisbon, Portugal. Epidemic History of Hepatitis B Virus.

11th IAS Conference on HIV Science(IAS 2021); 18th-21st July, 2021; Berlin, Germany; Epidemiological History of Hepatitis B Virus in Portugal and the Portuguese Role in Its Spreading Across the World

Para os meus filhos Sara e João

(To my children Sara and João)

ACKOWLEDGEMENTS

Perpétua Gomes, from Molecular Biology laboratory of Hospital Egas Moniz in Lisbon, thank you for the support you gave me specifically in this work: for the samples, for the sequencer and for letting me unwind in your always good company.

To Fátima, Quim, Isabel, Inês and Patrícia from "Perpétua's Laboratory" I want to thank you for having always received me so well in your group, being always available to help me with my sequencing reactions. It was a pleasure to work with all of you.

Rubén, the former master's student at "Perpétua's Laboratory", I thank you for all the help you've been giving me, even after moving yourself to British lands, with data, protocols, things I no longer remembered and wouldn't have easily accessed without your help. It was a pleasure working with you and I am happy with the path you are taking.

Marta Pingarilho and Victor Pimentel, from Ana Abecasis' group, thank you so much for your help and suggestions for my work, for your sympathy and for making me feel always welcome to your group.

To all colleagues from my current workplace, Faculty of Pharmacy of University of Lisbon, Inês Bártolo, Inês Moranguinho, Susana Tracana, Marta Calado, Claudia Palladino, Paloma Gonçalves, Pedro Gomes, João Perdigão, thank you so much for your company, help, our conversations, snacks, delicious desserts, sense and nonsense talks and above all for making me feel how is to work as a team again. And no one better than you to share my "suffering PhD" from time to time. We are all in the same boat and one by one we are docking at the port. Good luck to all! Always count on me!

To you, Nuno Taveira, are there any words? I feel like I owe you most of my professional life. Even what ended up going less well, had a big positive impact on what I know today and on who I am. And the thing is, you're always there when there's no one left, you show up and don't let everything fall apart. Thank you for all what you have done when nothing of this had to do with you, thank you for the opportunities, for the help with your ideas, suggestions and super-mega-quick revisions of the thesis and papers. I admire so much all that scientific knowledge that live inside you, that eternal passion for science so characteristic of you and so unique. Thank you so much for your presence and your friendship. I will always be there for you!

To the Marcelinos, that extended family that so many years ago welcomed me with open arms. Parents' in-law, sisters-in-law, brother-in-law, nephews, nieces and even cousins, aunts and uncles, thank you for all your support and friendship throughout this life. Especially in this period of my PhD thesis, I must thank particularly to my sisters-in-law Nela and Nanda for your support in so many things…not only because the rides to work but also the way you contributed to help me focus on the thesis in less easy times and regain inspiration to write. And to you, Andreia, not only for the first review of my paper, but for becoming the other half of me in science, the part that I decided to leave behind me, but that I am glad to see in you every day. You had the courage to look for consistence in your professional life abroad, far away it is true, an ocean away, but it will serve you to have a strong professional life here when you return to us. For me, the 10-year-old girl I once knew has become a model of courage and passion for science to follow.

To my family ♥, that is present since always, my parents and my sister, and since a long time too, my brother-in law and nephews, thank you for being who you are, for your presence in my life. To my parents, Carlos and Irene, what can I thank you for, that I haven't already said in the two previous theses? Thank you for the high standards set that you have always instilled me, for your patience with these last years of my almost constant absence. This took me more time and it was more troubled than I expected, but now I see the light in the end of the tunnel! Thank you for your patient support.

Vera and Bruno, my sister and my brother-in-law, thanks for your support too, for the joy at your home, for the nonsense conversations, Vera, that in fact always make sense for me. I think this is some kind of "mind-sisters" connection, but your nonsenses are kind of relaxing! Bruno, thank you so much for the precious help to finish this path, you managed to surprise me, and I even don't know why...after all, you are always so dedicated to the family that I couldn't wait for anything else.

Miguel and André, my sweet and adorable little "demons", thank you for your contagious happiness and inexhaustible energy that make family meetings so happy (and João so tired!) and that worked as a source of relaxation for me at various times! You are fantastic nephews! Miguel, less broken heads is just an idea for the future, ok?? That doesn't help to relax...

To my dear *everyday-forever-home-family* ♥, Zé, Sara and João, you are everything to me. Zé, thank you for being by my side, for starting all this, for believing in me, for trying to keep me from getting discouraged, for trying to take worries away from me even though sometimes I prefer to worry, for being able to put up with my bad temper on my most difficult days, and for replacing me in so many things over this last year so I could have more time to finish this thesis. Thanks for pushing me until I reach the end…I wouldn't do this without your strength to complete me whenever I felt mine slipping away, your always positive attitude whenever my despair gained ground in me. This is just one example of why we've been together for so long, what you have and I don't have, and vice versa makes us a whole, completes us and makes us move forward through the ups and downs of life. Thank you for being who you are and always being there for me!

Sara and João, my dear children, it was you who suffered the most from the consequences of this turbulent last year, because I was not as present as usual. Sometimes a little bit more forgotten, sometimes distant, sometimes leaving too early and coming back too late, without going out on the weekend to help you unwind and enjoy your free time in family and without giving you the usual support in your studies. And yet, with ups and downs, you've risen to the occasion. Sara, look at what you've accomplished this year for yourself! Never say you can't after this! João, this year I realized that you can do things for yourself with an excellent relaxed attitude and you do it so well! I am so proud of your success children and how much you've grown! I will not forget your stimulating support: "You can do it, Mom!", "Go and write that thesis!" and sometimes…"Is there still a long way to go?", or..."Have you still only written 4300 words??"... well, everything was support, everything was encouragement. We are going to achieve our things like this, the four of us together, always, no matter what happens.

…I know, I know, I will not forget Christie, children! It arrived in the perfect moment, and it is a good and true companion, very supportive although it doesn't talk too much…the positive side is that it never asked me "How many pages are still left?" …good girl!

RESUMO

Em 2019, a Organização Mundial de Saúde estimou 354 milhões de pessoas com hepatite B e C crónica a nível mundial. Em Portugal, o mais recente inquérito transversal a nível nacional parece revelar uma diminuição das prevalências de VHB e VHC anteriormente estimadas. A informação sobre a evolução da diversidade viral continua limitada, desconhecendo-se se a epidemiologia molecular das infeções por VHB e VHC tem vindo a ser alterada em Portugal.

O objetivo deste estudo foi investigar a história evolutiva de ambos os vírus e a sua filogeografia em Portugal.

Os ácidos nucleicos virais foram extraídos do plasma de doentes portugueses com VHB e VHC, os genes *pol* do VHB e *NS5A* e *NS5B* do VHC foram amplificados e sequenciados. A reconstrução de árvores filogenéticas de máxima verossimilhança no software IQtree v1.6.11 foi usada para genotipar as estirpes virais. A história epidémica dos subtipos do VHC em Portugal foi reconstruída através de métodos Bayesianos, conforme implementados no software BEAST v1.10.4, tal como a investigação da origem e das rotas de dispersão de ambos os vírus. As resistências aos antivirais nos dois clades do subtipo 1a do VHC foram analisadas online no Geno2Pheno [HCV] 0.92.

Os nossos dados revelaram que o subgenótipo D4 do VHB foi o primeiro a ser introduzido em Portugal cerca de 1857, seguido pelo D3 e A2 algumas décadas mais tarde. Os genótipos E e A1 do VHB foram introduzidos posteriormente, quase em simultâneo. Os nossos resultados revelaram um papel muito importante de Portugal na exportação de D4 e A2 para o Brasil e Cabo Verde, respetivamente, no início do século XX.

Em Portugal foram identificados subtipos distintos de VHC que entraram no país ao longo do tempo: subtipo 1b (1930-1960), subtipos 3a (1960s) e 1a (1980s), possivelmente associados a transfusões de sangue contaminado, ao início do uso de drogas intravenosas e ao seu uso generalizado, respetivamente. Os subtipos 4a e 4d, emergiram mais recentemente, possivelmente com o ressurgimento do uso de opiáceos. O subtipo 1a é claramente o mais frequente em Portugal apresentando dois clades diferentes (I e II) a circular na população, que possivelmente possuem vias de transmissão diferentes. Os primeiros países a introduzir os clades I e II em Portugal foram os Estados Unidos da América (1965) e Espanha (1955), respetivamente. Dois subclades, classificados como X e Y, foram identificados entre as estirpes do clade I. As RAS basais no gene *NS5A* foram encontradas principalmente nas estirpes do clade I/subclade Y, sendo a mutação mais frequente a L31M, que se revelou ausente nas estirpes do clade I/subclade X e do clade II.

Este trabalho permitiu conhecer a história epidemiológica do VHB e VHC em Portugal, mostrando que Portugal teve um papel importante na dispersão global do VHB e fornecendo novos conhecimentos sobre a epidemiologia molecular, origem e dinâmica de dispersão do VHC em Portugal. Também indicou algumas estirpes virais de VHC como mais propensas a adquirir RAS e que a resistência aos antivirais deve ser investigada no contexto dos clades/subclades do subtipo 1a.

Palavras-chave: Hepatites Virais; Portugal; Evolução Viral; Reconstrução Bayesiana Espaço-Temporal; Filogeografia

.

ABSTRACT

In 2019, the World Health Organization estimated 354 million people with chronic hepatitis B and C worldwide. In Portugal, the most recent national cross-sectional survey indicates a decrease in previously estimated HBV and HCV prevalence. Information on the evolution of viral diversity remains limited, and it is not known whether the molecular epidemiological profile of HBV and HCV infections has been changing in Portugal.

The aim of this study was to investigate the molecular epidemiology and evolutionary history of both viruses and their phylogeography in Portugal.

Viral nucleic acids were extracted from the plasma of Portuguese patients infected with HBV and HCV, the *pol* gene of HBV and *NS5A* and *NS5B* genes of HCV were amplified and sequenced. The reconstruction of maximum likelihood phylogenetic trees in the IQtree v1.6.11 software was used for genotyping purposes. The epidemic history of HCV subtypes in Portugal was reconstructed using Bayesian methods, as implemented in the BEAST v1.10.4 software, as in the investigation of the origin and routes of spread of both viruses. Antiviral drug resistance in the two HCV subtype 1a clades were analyzed online in Geno2Pheno [HCV] 0.92.

Our data indicated that the D4 subgenotype of HBV was the first to be introduced in Portugal around 1857, followed by D3 and A2 a few decades later. HBV genotypes E and A1 were introduced later, almost simultaneously. Our results revealed a very important role of Portugal in the exportation of D4 and A2 to Brazil and Cape Verde, respectively, in the beginning of the 20th century.

Distinct HCV subtypes that entered Portugal over time were identified: subtype 1b (1930-1960), subtypes 3a (1960s) and 1a (1980s), possibly associated with transfusions of contaminated blood, with the beginning of intravenous drug use and its widespread use, respectively. Subtypes 4a and 4d have emerged more recently, possibly with the resurgence of opiate use. Subtype 1a is clearly the most frequent in Portugal with two different clades (I and II) circulating in the population, which possibly have different transmission routes. The first countries to introduce clades I and II in Portugal were the United States of America Spain (1965) and Spain (1955), respectively. Two subclades, classified as X and Y, were identified among the clade I strains. The basal RAS in the NS5A gene were found mainly in the clade I/subclade Y strains, the most frequent mutation being L31M, which was absent in the strains of clade I/subclade X and clade II.

This work allowed us to understand the epidemiological history of HBV and HCV in Portugal, showing that Portugal played an important role in the global spread of HBV and providing new knowledge about the molecular epidemiology, origin and dispersion dynamics of HCV in Portugal. It also highlighted that some viral strains of HCV may be more likely to acquire RAS and that antiviral resistance should be further investigated in the context of subtype 1a clades/subclades.

Keywords: Viral Hepatitis, Portugal, Viral Evolution, Bayesian Spatiotemporal Reconstruction, Phylogeography

TABLE OF CONTENTS

CHAPTER 3
RESULTS   **77**

CHAPTER 4
DISCUSSION AND CONCLUSIONS   **111**

TABLE INDEX

LIST OF ABBREVIATIONS

Aa	Amino acids
aHBV	ancient Hepatitis B virus
B.C.	Before Christ
BCE	Before Common Era
BI	Bayesian Inference
BLAST	Basic Local Alignment Search Tool
bp	Base pair
BSP	Bayesian skyline plot
C	Core
cccDNA	Covalently Closed Circular DNA
CE	Common Era
CI	Confidence Interval
CHLO	Centro Hospitalar de Lisboa Ocidental
CLDN1	Claudin-1 (tight junction proteins)
CRFs	Circulating Recombinant Forms
DAA	Direct Acting Antiviral
DGS	Direção Geral da Saúde
DMSO	Dimethyl Sulfoxide
DNA	Deoxynucleic Acid
dNTPs	Deoxynucleotides triphosphate
dsDNA	Double stranded DNA
DTT	Dithiothreitol
EASL	European Association for the Study of the Liver
ECDC	European Center for Diseases Control
EMCDDA	European Monitoring Centre for Drugs and Drug Addiction
ER	Endoplasmic Reticulum
ESS	Effective Sample Size

List of abbreviations

EU/EEA	European Union/European Economic Area
FW primer	Forward primer
GAG	Glycoaminoglycans
GBD	Global Burden of Disease
GTR	General-Time-Reversible
HAV	Hepatitis A virus
HBcAg	Hepatitis B c antigen
HbeAg	Hepatitis B e antigen
HBsAg	Hepatitis B surface antigen
HBx	Hepatitis B x (protein)
HBV	Hepatitis B virus
HBV/A	Hepatitis B virus genotype A
HBV/A1	Hepatitis B virus subgenotype A1
HBV/A2	Hepatitis B virus subgenotype A2
HBV/A3	Hepatitis B virus subgenotype A3
HBV/B	Hepatitis B virus genotype B
HBV/B2	Hepatitis B virus subgenotype B2
HBV/C	Hepatitis B virus genotype C
HBV/C1	Hepatitis B virus subgenotype C1
HBV/D	Hepatitis B virus genotype D
HBV/D1	Hepatitis B virus subgenotype D1
HBV/D2	Hepatitis B virus subgenotype D2
HBV/D3	Hepatitis B virus subgenotype D3
HBV/D4	Hepatitis B virus subgenotype D4
HBV/E	Hepatitis B virus genotype E
HBV/F	Hepatitis B virus genotype F
HBV/F2	Hepatitis B virus subgenotype F2
HBV/F2a	Hepatitis B virus subgenotype F2a
HBV/F1b	Hepatitis B virus subgenotype F1b
HBV/F4	Hepatitis B virus subgenotype F4
HBV/G	HBV genotype G
HBV/H	HBV genotype H

HBV/I	HBV genotype I
HBsAg	HBs antigen
HCC	Hepatocellular carcinoma
HCV	Hepatitis C virus
HCV/1	HCV genotype 1
HCV/1a	HCV genotype 1/subtype a
HCV/1b	HCV genotype 1/subtype b
HCV/1g	HCV genotype 1/subtype g
HCV/2	HCV genotype 2
HCV/2a	HCV genotype 2/subtype a
HCV/2c	HCV genotype 2/subtype c
HCV/3	HCV genotype 3
HCV/3a	HCV genotype 3/subtype a
HCV/4	HCV genotype 4
HCV/4a	HCV genotype 4/Subtype a
HCV/4b	HCV genotype 4/Subtype b
HCV/4d	HCV genotype 4/subtype d
HCV/4f	HCV genotype 4/Subtype f
HCV/4k	HCV genotype 4/Subtype k
HCV/6	HCV genotype 6
HDV	Hepatitis D virus
HEV	Hepatitis E virus
HIV	Human Immunodeficiency Virus
HIV-1	Human Immunodeficiency Virus type 1
HPD	Highest Posterior Density
HPD 95%	Highest Posterior Density Interval
HSPGs	Heparan sulphate proteoglycans
ICTV	International Committee on Taxonomy of Viruses
IQR	Interquartile Range
IRES	Internal Ribosomal Entry Site
Kb	Kilobase
kDa	Kilo Dalton

List of abbreviations

KU Leuven	Katholieke Universiteit Leuven
LDLr	Low density lipoprotein receptors
LHB	Large Hepatitis B (protein)
LiPA	Line Probe Assay
LVP	Lipoviral particles
MCC	Maximum clade credibility
MCMC	Markov Chain Monte Carlo
MERS-CoV	Middle East Respiratory Syndrome Coronavirus
MHB	Medium Hepatitis B (protein)
Min	Minutes
ML	Maximum Likelihood
MRCA	Most Recent Common Ancestor
mRNA	Messenger RNA
MSA	Multiple Sequence Alignment
MSM	Men who have Sex with Men
NCBI	National Centre for Biotechnology Information
Ne	Effective Population Size
NHP	Non-Human Primates
NJ	Neighbor-Joining
NTCP	Na(+)-taurocholate co-transporting polypeptide
OCLN	Ocludin
ORF	Open Reading Frame
OST	Opioid Substitution Therapy
P	Pol
PCR	Polymerase Chain Reaction
pgRNA	pre-genomic RNA
pRT	Para-Retroviruses
PSP	Posterior State Probability
PT	Portuguese
PWID	People Who Inject Drugs
rcDNA	Relaxed circular DNA
RAS	Resistance Associated Substitution

RNA	Ribonucleic Acid
RT	Reverse Transcriptase
RT-PCR	Reverse Transcription followed by Polymerase Chain Reaction
RV primer	Reverse primer
S	Surface
SARS-CoV2	Severe Acute Respiratory Syndrome Coronavirus 2
SD	Standard Deviation
Sec	Seconds
SHB	Small Hepatitis B (protein)
SICAD	Serviço de Intervenção nos Comportamentos Aditivos e nas Dependências
SRBI	Scavenger receptors class B type I
ssRNA	Single stranded RNA
s/n/c	Substitutions per nucleotide per cell infection
s/s/y	Substitution/Site/year
SVR	Sustained Virological Response
TAE	Tris-Acetate-EDTA
tMRCA	time for the Most Recent Common Ancestor
UCLD	Uncorrelated Lognormal relaxed molecular clock
USA	United States of America
UTRs	Untranslated Regions
VHB	Vírus da Hepatite B
VHC	Vírus da Hepatite C
Vs	Versus
WHO	World Health Organization

INTRODUCTION

Part I

Viral Hepatitis as a major public health problem

1.1. General aspects of Hepatitis

Viral hepatitis is characterized by inflammation of the liver resulting from infection with a virus. Although patients may remain asymptomatic at an early stage, over time, progression to liver fibrosis, cirrhosis and/or hepatocellular carcinoma (HCC) may occur. This disease can be caused by five viruses: Hepatitis A virus (HAV), Hepatitis B virus (HBV), Hepatitis C virus (HCV), Hepatitis D virus (HDV) and Hepatitis E virus (HEV), that differ in several ways such as routes of transmission, severity of the illness, geographic distribution and prevention methods (1, 2). In 2017, the World Health Organization (WHO) published a global hepatitis report describing, for the first time, the global and regional estimates of viral hepatitis prevalence up to the year 2015, setting the baseline for tracking the progress of a newly implemented global strategy to achieve the elimination of these public health threats until 2030 (3). This report focused mainly on Hepatitis B and C because they lead to chronic disease in hundreds of millions of people and, together, they have been responsible for 96% of mortality associated with hepatitis.

1.2. Global perspective of HBV and HCV infections

In 2015, it was estimated that around 328 million people worldwide were living with hepatitis B (~257 million) and hepatitis C (~71 million). It was also estimated that there were 1.34 million deaths caused by viral hepatitis in 2015, a similar number of deaths as those caused by tuberculosis and a higher number than those caused by Human Immunodeficiency Virus (HIV) (Table 1.1). The mortality due to viral hepatitis has been raising because, for most of the infected patients, testing and treatment remains unreachable (3). To comply with the plan to eliminate these pandemics by 2030, WHO established as primary goal a 30% reduction in new cases of chronic viral hepatitis B and C infections and 10% reduction in viral hepatitis B and C deaths by 2020 and a final goal of 90% reduction in new cases of chronic viral hepatitis B and C and 65% reduction in viral hepatitis B and C deaths by 2030 (4).

Table 1.1 Evolution of Hepatitis B and C infections, diagnosis, treatment, and deaths between 2015 and 2020.

	Hepatitis B		Hepatitis C	
Estimated number of infections by both viruses globally	328 million (2015)			
	354 million (2019)			
Estimated number of Chronic Infections globally	257 million (2015)	296 million (2019)	71 million (2015)	58 million (2019)
Percentage of diagnoses (among infected patients)	9% (2015)	10% (2020)	20% (2015)	21% (2020)
Effective treatment	Tenofovir		Direct Acting Antivirals	
Percentage of treatments (among diagnosed patients)	8% (2015)	22% (2020)	7.4% (2015)	62% (2020)
Total number of deaths	1.34 million (2015)			
	1.10 million (2020)			

To achieve these goals five main strategic points were set by WHO (3-5).

a) Information for focused action: developing a strong strategic information system to understand viral hepatitis epidemics and focus the response.

b) Interventions for impact: defining essential, high-impact interventions on the continuum of hepatitis services that should be included in health benefit packages.

c) Delivering for equity: To strengthen health and community systems to deliver high-quality services to achieve equitable coverage and maximum impact.

d) Financing for sustainability: proposing strategies to reduce costs, improve efficiencies and minimize the risk of financial hardship for those requiring the services.

e) Innovation for acceleration: promoting and embracing innovation to drive rapid progress.

In the report of 2017, some concern was expressed about the possibility of not fully meeting the goals set for 2020, mainly due to the inequalities in the strategy's implementation observed in different WHO regions, in different countries from the same region and in access to diagnosis and treatment These worries concerned not only the general population, but also specific key populations such as women in childbearing age from low-income countries (which are potential HBV transmitters to their babies), prisoners, men who have sex with men (MSM), persons who inject drugs (PWID) and immigrants or refugees (3, 6-9). In fact, the 2021 WHO update report indicates a higher total number of infections by both viruses (354 million) in 2019 (4). The data show that, although the percentage of HCV patients treated with effective direct acting antivirals (DAAs) increased from 7.4% in 2015 to 62% in 2020, with a consequent decrease in the number of chronic infections, the reduction in mortality from both hepatitis was not significant (Table 1.1). Access to prevention, harm reduction and health-care services for some populations is still largely insufficient, and persistent stigma, inequalities, criminalization, and other socio-structural barriers are preventing response efforts from reaching the people who need them the most (4). For these reasons, the incidence and mortality rates of both viruses, still present a large variation between different regions of the world (Figure 1.1).

As for HBV, until 2013, the highest prevalence had been detected in the WHO African region, followed by the Western Pacific region (10). In 2020, the prevalence in Africa still remained the highest, but Western Pacific region had already made major progress in expanding access to services for HBV and also to HCV, largely because of domestic funding and substantial drug price reductions (4). And although Western Pacific still have the highest mortality from HBV in the world, the number of new infections is now higher not only in Africa but also in Southeast Asia (Figure 1.1). The region of Americas and Europe seems to be less affected by HBV when compared to the rest of the world.

Most of the burden of disease infection comes from HBV acquisition before the age of 5 through mother-to-child transmission. Women of childbearing age can potentially transmit HBV to their babies, especially in low-income countries where disease information, antenatal care, diagnosis access and treatment are limited (3). Despite available and affordable HBV medicines (Table 1.1), treatment seems to be progressing more slowly than expected (4). HCV, on the other hand, is prevalent worldwide, but the

highest prevalence registered in 2015 was in WHO Eastern Mediterranean Region followed by the European Region (3).

GLOBAL
Hepatitis B
New Infection: **1 500 000**
[1 100 000–2 600 000]
Deaths: **820 000**
[450 000–950 000]
Hepatitis C
New Infection: **1 500 000**
[1 300 000–1 800 000]
Deaths: **290 000**
[230 000–580 000]

REGION OF THE AMERICAS
Hepatitis B
New infections: 10 000
[5 100–26 000]
Deaths: 15 000
[8 500–23 000]
Hepatitis C
New infections: 67 000
[63 000–73 000]
Deaths: 31 000
[19 000–84 000]

EUROPEAN REGION
Hepatitis B
New infections: 19 000
[9 400–38 000]
Deaths: 43 000
[34 000–51 000]
Hepatitis C
New infections: 300 000
[240 000–320 000]
Deaths: 64 000
[39 000–72 000]

WESTERN PACIFIC REGION
Hepatitis B
New infections: 140 000
[96 000–210 000]
Deaths: 470 000
[200 000–490 000]
Hepatitis C
New infections: 230 000
[220 000–260 000]
Deaths: 77 000
[77 000–140 000]

WHO REGIONS
- African Region
- Region of the Americas
- South-East Asia Region
- European Region
- Eastern Mediterranean Region
- Western Pacific Region
- Not applicable

AFRICAN REGION
Hepatitis B
New infections: 990 000
[660 000–1 600 000]
Deaths: 80 000
[47 000–110 000]
Hepatitis C
New infections: 210 000
[150 000–370 000]
Deaths: 45 000
[23 000–72 000]

EASTERN MEDITERRANEAN REGION
Hepatitis B
New infections: 100 000
[79 000–140 000]
Deaths: 33 000
[26 000–60 000]
Hepatitis C
New infections: 470 000
[240 000–520 000]
Deaths: 31 000
[31 000–74 000]

SOUTH-EAST ASIA REGION
Hepatitis B
New infections: 260 000
[180 000–590 000]
Deaths: 180 000
[140 000–300 000]
Hepatitis C
New infections: 230 000
[200 000–430 000]
Deaths: 38 000
[37 000–130 000]

Sources: WHO, 2021

Figure 1.1 | Hepatitis B and C new infections and mortality by WHO region, 2019. The highest HBV and HCV incidences are registered in African and Eastern Mediterranean WHO regions, respectively. As for the mortality, Western Pacific region has the highest rate for both viruses. (Image source: Global progress report on HIV, viral hepatitis and sexually transmitted infections, WHO 2021; licensed under CC BY-NC-SA 3.0 IGO).

1.2.1. Main progresses on reduction of HBV and HCV transmission

1.2.1.1. Improvement in HBV birth dose coverage

Mother-to-child transmission is the main route of HBV transmission. Although vaccine is available with a global coverage that reached 85% among children less than 5 years old (Table 1.2), the initial birth dose is not always administered at the correct moment (3, 11, 12).

Table 1.2 Progresses on reduction of HBV and HCV transmission.

	Hepatitis B			Hepatitis C
Main transmission routes	Mother to child transmission			Injection drug use; Unsafe health-care practices; MSM co-infected with HIV
Global coverage with the complete immunization schedule	84% in infancy – 2015	85% in infancy - 2019		No vaccine available
Strategies and results in incidence reduction	In children less than 5 years old			Better health care practices and blood donor testing (second half of the 20th century). Needles exchanging campaigns for PWID.
	4.7% (pre-vaccination era: 1980s-2000s	1.3% (2015)	< 1% (2020)	
Main contributors to unsuccessful control of the pandemics	Initial birth dose vaccination is still low at 39% (2015) Initial birth dose vaccination is still low at 43% (2019)			5% of health-care-related injections or blood transfusions remained unsafe (lead to 1.75 million new HCV infections in 2015). 3.9% reuse of injection equipment (2017).
	Other prevention interventions are available but insufficiently implemented, such as diagnosis tests and treatment that do not reach every region in the world equitably.			

Birth dose administration after delivery is crucial to prevent transmission of HBV from an infected mother to her child during delivery. In 2015, the global coverage for this dose reached only 39%, rising, after 4 years, to 43% (4). However, there are important regional discrepancies, for instance, WHO African region only achieved 6% of HBV birth dose coverage (4, 13). Despite this, there was a considerable reduction in HBV incidence in children younger than 5 years old, since the pre-vaccination era (4.7%) until 2020 (<1%) (Table 1.2).

1.2.1.2. Availability of HCV diagnostic tests and treatment

In Eastern Mediterranean region, HCV transmission seems to be related mainly to unsafe health-care practices, including unsafe healthcare injections (14). To reverse this scenario,

by the end of 2020 the Eastern Mediterranean Region has been leading the way in scaling up testing and treatment for HCV, accounting for 37% of the total number of diagnosed patients globally and 52% of the total number of treated patients globally (4). In Europe, the main route of HCV transmission is injection drug use but also seems to be common in MSM who are co-infected with HIV through sexual transmission (3, 15, 16). In 2020, the European region showed major advances in hepatitis program planning to reduce gaps in testing and treatment. Nonetheless, in the field of HCV prevention and harm reduction among PWID, WHO considers that the progress is still slow (4).

1.2.1.3. Access to DAA treatment in HCV infected patients

Incidence reduction of HCV cases are mostly focused on improving health care practices, blood donor testing and, more recently, also in needles exchange programs for PWID. The prevalence has been declining due to DAAs which are effective in eliminating the virus. It is important to note that the access to the DAAs is one of the best examples of the inequalities around the world. In 2015, out of the 5.5 million cumulative number of patients undergoing HCV therapies, only half a million received the new and effective DAA due to their exorbitant costs (3, 17). By that period, the number of new HCV infections in 2015 worldwide turned out to be higher than the number of patients who started these new treatments (3). Some countries face an economical funding problem given that governments only support the treatment of part of the infected population while the remaining patients must appeal to health security companies or even pay their own complete treatment regimen (18). The production of DAAs generics allowed for a reduction in the prices of these drugs, but not all countries have the same access to generics. In India only Sofosbuvir has a more affordable price which is why no other DAA is used for the treatment of HCV (19). However, the most recent data, resulting from a re-evaluation between 2019 and 2020, indicate that one of the aspects that experienced a promising improvement was the increase in the number of HCV patients who received DAA treatment, around 9.4 million, which reveals an increase of almost 10 times compared to 2015 (4). Price reductions have made hepatitis C treatment an affordable high-impact intervention, but to achieve the 2030 goals an additional effort is needed to increase even more the number of treated persons. Improving access to HCV treatment has led to a dramatic decrease in the number of deaths from HCV. However,

the total number of deaths from HCV and HBV remains high, at around 1.1 million deaths in 2020 (Table 1.1).

In summary, for HBV and HCV elimination programs to meet their goals by 2030, it is imperative to accelerate progress, address specific gaps in implementation, and bring innovation to scale up the response to these two pandemics (4, 20). The main priorities identified are:

a) implementation of HBV vaccine-based strategies to prevent perinatal transmission.
b) safe health care practices to prevent hepatitis C transmission through contaminated blood and injecting paraphernalia.
c) safe injection practices and HCV treatment for PWID.
d) Expanding diagnosis and treatment for HBV and HCV to all those who need them, through committed political leadership, and a reduction in the prices of essential medicines and diagnostics.

1.3. The situation in European Union and European Economic Area

The European Union/ European Economic Area (EU/EEA) is considered a low prevalence area for HBV and HCV, having an estimated total of 4.7 million chronic HBV cases and 3.9 million chronic HCV cases (21). The countries of the EU/EEA are committed to implement efforts to achieve WHO global targets for viral hepatitis for 2020 and 2030. However, there are many countries that do not report data, making it difficult to draw conclusions about progress towards the targets (Table 1.3) (21-24).

Table 1.3 Countries progress on the hepatitis continuum of care and on the prevention of new hepatitis infections in the EU/EEA, 2017.

Country	Hepatitis B					Hepatitis C			
	HBV vaccine coverage	% diagnosis	% treated	% with viral suppression	Syringes distributed /PWID	OST coverage	% diagnosis	% treated	% of cure
TARGET	95%	50%	75%	90%	200	40%	50%	75%	90%
Austria	90	No data	61.0	No data	No data	50	35.3	27.7	No data
Belgium	97	No data	No data	No data	50	No data	No data	No data	No data
Bulgaria	92	2.4	100.0	28.6	No data	No data	4.1	35.7	100.0
Croatia	94	No data	No data	No data	192	54	24.8	5.8	99.1
Cyprus	97	No data	No data	No data	1	18	No data	No data	No data
Czechia	94	No data	No data	No data	147	38	No data	No data	No data
Denmark	NA*	71.8	No data	No data	No data	No data	44.3	No data	No data
Estonia	92	11.0	No data	No data	232	No data	29.1	No data	No data
Finland	NA*	No data	No data	No data	373	No data	No data	No data	No data
France	90	17.5	No data	No data	109	85	80.6	18.1	90.0
Germany	87	No data	No data	No data	No data	54	No data	No data	No data
Greece	96	35.0	No data	No data	76	65	20.0	No data	No data
Hungary	No data	No data	No data	No data	21	No data	25.9	8.4	92.7
Iceland	NA*	No data	No data	No data	NA**	NA**	77.2	100.0	95.0
Ireland	95	57.1	No data	No data	No data	54	63.4	10.1	98.0
Italy	94	No data	No data	No data	No data	30	No data	No data	No data
Latvia	98	24.0	No data	No data	108	9	96.8	5.8	91.2

Country									
Leichtenstein	NA*	No data	No data	No data	NA**	NA**	No data	No data	No data
Lithuania	94	No data	No data	No data	28	15	No data	No data	No data
Luxembourg	94	No data	No data	No data	305	66	No data	No data	No data
Malta	88	No data	No data	No data	No data	72	No data	No data	No data
Netherlands	92	51.5	22.2	No data	No data	No data	No data	No data	No data
Norway	No data	No data	4.3	No data	332	No data	No data	No data	No data
Poland	93	47.3	No data	No data	No data	18	24.9	No data	No data
Portugal	98	No data	No data	No data	108	45	No data	No data	96.8
Romania	92	7.1	21.7	85.8	No data	8	7.6	42.0	100.0
Slovakia	96	14.6	No data	No data	No data	No data	92.7	No data	No data
Slovenia	89	No data	28.1	100.0	No data	62	No data	13.6	95.0
Spain	93	No data	No data	No data	119	No data	No data	35.6	100.5
Sweden	76	No data	No data	No data	No data	No data	No data	No data	No data
United Kingdom	No Data	59.3	No data	No data	No data	57	53.5	16.1	96.0

*Not applicable because country did not have a universal childhood HBV vaccination program. ** Not applicable because country was not included in data collection efforts. PWID, People Who Inject Drugs; OST, Opioid Substitution Therapy

Source: Table A from the technical report of European Center for Diseases Control (ECDC), "The sustainable development goals and hepatitis B and C in the EU/EEA", 2021 (24), which was based on data from the technical report of ECDC, "Monitoring the responses to hepatitis B and C epidemics in the EU/EEA Member States", 2019 (22); ECDC copyright protected but reproduction is authorized provided the source is cited.

1.3.1. The Continuum of Care

The targets for diagnosis and treatment of HBV and HCV form a continuum of care. The European action plan to reach WHO targets for 2020 (5) comprised:

- achieving the diagnosis of 50% of those with chronic HBV and HCV.
- starting treatment of 75% of diagnosed patients who are eligible for treatment.
- reaching viral suppression (for HBV) or a sustained virological response (for HCV) in 90% of those who receive treatment.

To progress towards those targets, EU/EEA established that the most critical steps (24) are:

- 95% coverage of the complete immunization schedule (three doses of HBV vaccine) in countries that implemented universal childhood vaccination.
- To distribute a minimum of 200 syringes per PWID per year (HCV prevention).
- To improve the availability of opioid substitution therapy (OST), making it available to at least 40% of opioid dependent PWID (HCV prevention).

1.3.1.1 HBV childhood vaccination coverage

As for the EU/EEA 2020 plan of 95% coverage with three doses of HBV vaccine in childhood, a total of 27 out of 31 EU/EEA countries recommended universal childhood vaccination against hepatitis B in 2017 (table 1.3) (24). Three countries did not have a national policy for universal vaccination (Denmark, Finland, and Iceland) and one country (Sweden) had regional implementation of universal hepatitis B vaccination as of 2017. Data on vaccine coverage in 2017 were reported from 24 countries. Of these, seven countries (29%) had reached the 2020 target of 95% childhood coverage: Latvia, Belgium, Cyprus, Greece, Slovakia, Ireland and Portugal (24).

1.3.1.2 Harm reduction among PWID

Data on coverage of prevention programs targeting PWID are lacking from half the countries in the EU/EEA (Table 1.3). Four out of 15 (29%) countries with data in 2017 had coverage of at least 200 syringes distributed per PWID per year: Estonia,

Luxembourg, Norway and Finland (25). Eighteen countries had estimates of the population of high-risk opioid users. Eleven of these countries (61%) reported achieving the 2020 target of at least 40% coverage of opioid substitution therapy: France, Austria, Germany, Croatia, Ireland, England, Slovenia, Greece, Luxembourg, Malta, and Portugal.

1.3.1.3. Diagnosis and treatment

The monitorization of continuum of care of HBV is limited by the lack of countries that report the estimated number of people living with the virus and the number of those who are diagnosed. By gathering data from the 12 countries reporting data on both parameters, there were an estimated 1 597 377 people with chronic HBV infection in EU/EEA, of whom 20.3% (range 2.4 – 71.8%) were reported to have been diagnosed (23). Four of the 12 countries had met or exceeded the diagnosis goal of 50% of those with chronic HBV: Denmark, Ireland, Netherlands, and United Kingdom. From those 12 countries, three reported data on the percentage of diagnosed persons who started HBV treatment (Table 1.3). Bulgaria seems to be the only one achieving the 2020 goal of 75% treatment among the HBV diagnosed persons. However, many cases diagnosed with chronic infection may not be eligible for treatment according to clinical guidelines and the data contributing to this estimate were not adjusted to account for these cases. No country reported data on both the number of diagnosed and treatment eligible individuals with chronic HBV infections and the number of those receiving treatment, so it was not possible to measure progress towards the treatment target for HBV (24).

As for HCV, there were an estimated 1 422 285 people with chronic HCV infection, of whom 26.8% (range 4.1 – 96.8%) have been diagnosed. Six of the 15 countries had met the goal of 50 % diagnosis of those with chronic HCV in 2017: France, Iceland, Ireland, Latvia, Slovakia, and United Kingdom (Table 1.3). More than half of the countries did not report data about this goal, but for those which have done it, 23.0% (range 5.8 – 100%) had been started on treatment. Iceland was the only country that achieved the 2020 target of 75% of those with diagnosed infections starting treatment in 2017 (24).

1.3.1.4. Viral suppression of HBV and viral elimination of HCV

For chronic HBV cases, only three countries reported data on the proportion of those treated who attained viral suppression. Slovenia was the only country of the three to reach the 2020 target of having 90% of treated patients with viral suppression in 2017 (22). For chronic hepatitis C cases, 12 countries reported data on the proportion of those treated who had a sustained virological response (SVR), or cure. Among the 12 countries, 92.4% of those treated achieved SVR. All 12 countries had reached the 2020 target of having 90% of treated patients with SVR in 2017: Bulgaria, Croatia, France, Hungary, Iceland, Ireland, Latvia, Romania, Slovenia, Spain, United Kingdom, and Portugal.

1.3.1.5. Mortality rate

The mortality proposed by WHO as an indicator of HBV and HCV progress is represented by the number of deaths by liver disease attributable to HBV and HCV infection, e.g. hepatocellular carcinoma (HCC) and chronic liver diseases, including cirrhosis (26).
The baseline mortality of EU/EEA, that can be attributed to HBV and HCV, was estimated in 2015 at 63 927 deaths (27) varying by region of Europe (Figure 1.2). To reach the mortality WHO global target, a total of 41 553 deaths in EU/EEA should be prevented by 2030 (65% of the 2015 baseline) (27).

In summary, although reliable data on most indicators of HBV and HCV progress are lacking, the data available shows that most EU/EEA countries are far from reaching the WHO elimination targets for hepatitis B and C (24). It is, however, evident that the greatest progress has been achieved among those with chronic hepatitis C who received treatment and who achieved SVR, or cure. All countries reporting data met the 2020 target in the treatment of patients infected with HCV with DAA.
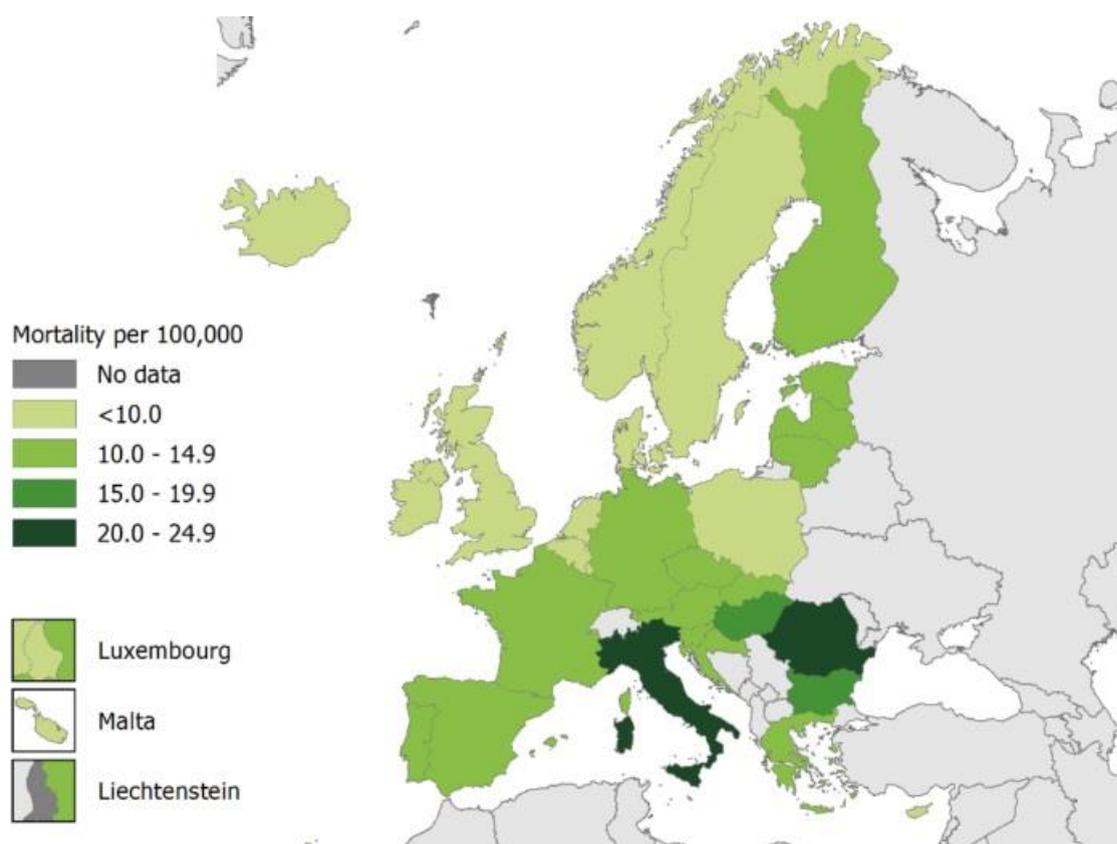
Figure 1.2 | Mortality attributable to HBV and HCV by EU/EEA country in 2015. Number of deaths per 100 000 population from hepatocellular carcinoma and cirrhosis and other chronic liver diseases combined attributable to HBV and HCV (Image source: The sustainable development goals and hepatitis B and C in the EU/EEA, ECDC 2021; ECDC copyright protected but reproduction is authorized provided the source is cited).

1.4. Epidemiology of HBV and HCV in Portugal

1.4.1. Hepatitis C

Despite new national initiatives and efforts needed to achieve all the 2030 WHO targets, there are some important improvements and visible progress in the general framework of the HCV elimination in Portugal.

1.4.1.1 HCV seroprevalence in general population

In Portugal, only a few national studies have estimated the prevalence of HCV infection. Studies carried out between 2012 and 2016 in the general population reveal a possible decrease in the prevalence of HCV in Portugal in recent years (28-31). A nationwide

cross-sectional survey enrolling more than 1600 adults selected randomly from the lists of primary health care settings between 2012 and 2014, reported an HCV prevalence of 0.54% [Confidence interval (CI): 0.2 – 0.9] (28). In addition, the results found in the last National Serological Survey in 2015-2016, revealed an even lower seroprevalence for HCV: 0.3% (CI: 0.1 - 0.6), weighted for the resident Portuguese population on the continent and islands (29). Despite a possible underrepresentation of high-risk groups in these recent surveys that may explain such a remarkably low HCV prevalence (28, 29), it still seems to be lower than previous estimates of 1 – 1.5% (30, 31).

1.4.1.2 HCV seroprevalence among PWID and harm reduction

According the European Monitoring Centre for Drugs and Drug Addiction (EMCDDA) and Serviço de Intervenção nos Comportamentos Aditivos e nas Dependências (SICAD), the most recent estimated population size of intravenous drug users in Portugal is 13162 persons (22, 25, 32), with diagnosis test results in PWID showing an HCV seroprevalence of 81.5% (Figure 1.3).
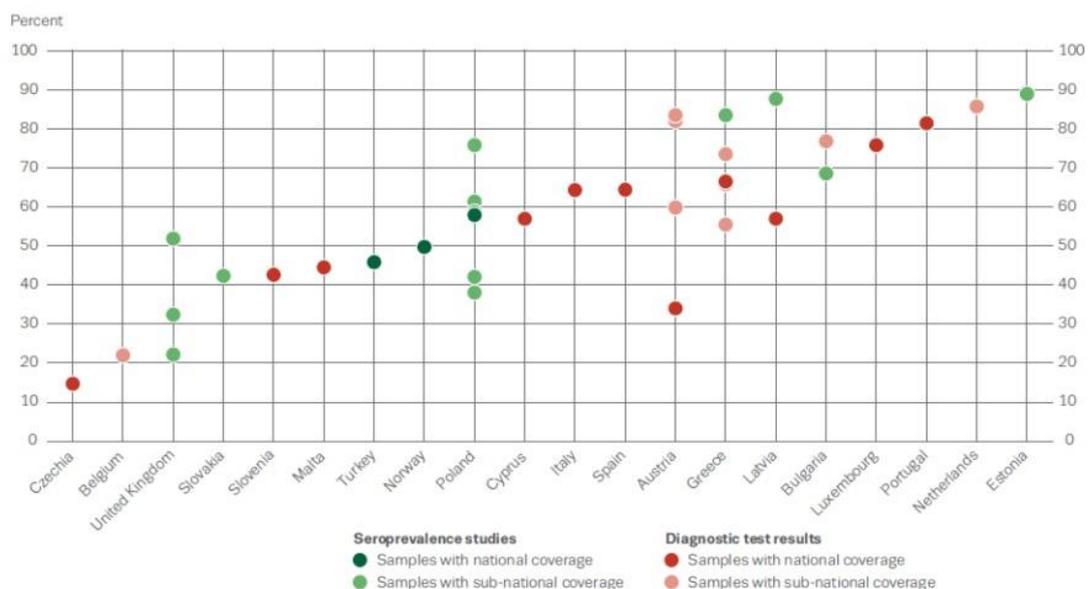
Figure 1.3 | HCV antibody prevalence (%) among PWID, 2016-2017. Results were obtained from seroprevalence studies and diagnostic tests with national and sub-national coverage. Data from Portugal were reported according to diagnostic tests with national coverage in 2017. (Image source: Monitoring the elimination of viral hepatitis as a public health threat among people who inject drugs in Europe: the elimination barometer, EMCDDA 2019; EMCDDA copyright protected but reproduction is authorized provided the source is cited).

Among the total number of drug addicts who resorted to treatment (injectors and non-injectors), HCV seroprevalence ranged between 10% and 59% depending on the subgroup they belong to [withdrawal units (52%), therapeutic communities (25%), outpatient clinics (59%), new clients undergoing treatment (10%) and clients readmitted to treatment (50%)]. If from this group only those who are drug injectors are considered the seroprevalence is much (between 54% and 89%) (32).

Despite a general decrease in heroin consumption in Portugal in comparison with drugs such as cocaine and cannabis (32), the most reported primarily injected opiate among drug treatment entrants of 2017 was still heroin, with 7.9 % of injectors reporting to have recently shared used needles/syringes (25). Portugal has adopted a national hepatitis control policy that covers PWID and includes the distribution of sterile syringes/needles and OST. In 2017, Portugal distributed 1 421 666 sterile syringes/needles, which resulted in 108 per PWID per year (25) a number that still did not achieve the 2020 WHO target of a minimum of 200 syringes per PWID per year (Figure 1.4) (24).
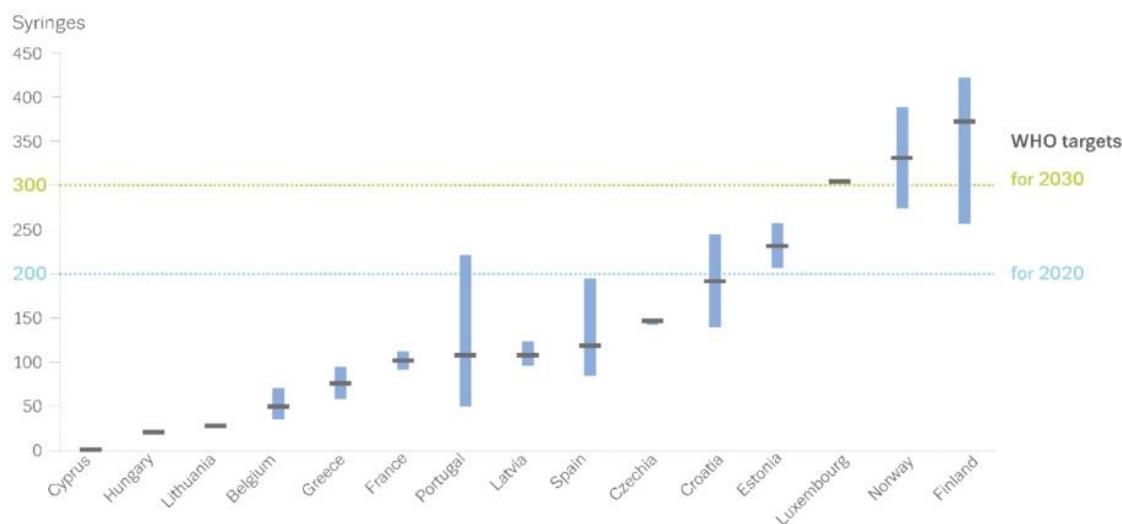
Figure 1.4 | Estimated number of syringes provided PWID in EU Member States and Norway, 2017.
(Image source: Monitoring the elimination of viral hepatitis as a public health threat among people who inject drugs in Europe: the elimination barometer, EMCDDA 2019; EMCDDA copyright protected but reproduction is authorized provided the source is cited).

Portugal turned OST available to 45% of opioid dependent PWID, (24), becoming one of the eleven countries with a coverage of OST superior to 40% as planned by WHO (Figure 1.5).
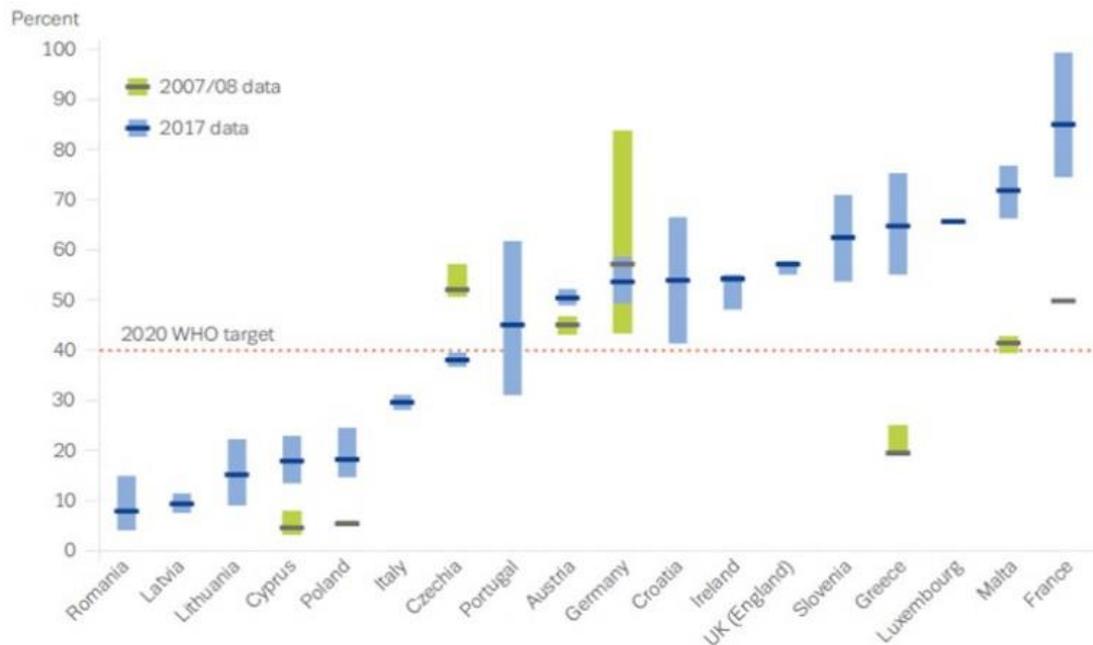
Figure 1.5 | Coverage of OST in 2017 or most recent year, and the same data in 2007 or 2008. Graphic shows percentage of estimated high-risk opioid users receiving treatment and the respective uncertainty interval. OST, Opioid substitution therapy. (Image source: Monitoring the elimination of viral hepatitis as a public health threat among people who inject drugs in Europe: the elimination barometer, EMCDDA 2019; EMCDDA copyright protected but reproduction is authorized provided the source is cited).

In the context of the prison population, where intravenous drug use is of common practice, the seroprevalence of Hepatitis C was 43% in 2019, the lowest in the last five years, decreasing for the second consecutive year and being lower than that of most groups of drug users undergoing treatment in free environment (32). In Portugal, a recent innovative strategy for micro-elimination was adopted in prisons, with DAA therapeutic schemes without Ribavirin (52% elbasvir/grazoprevir, 30% sofosbuvir/velpatasvir and 18% sofosbuvir/ledipasvir) achieving a remarkable success rate of 100% of sustained virologic response (33).

1.4.1.3 HCV treatment and sustained virological response

To comply with the global commitment for hepatitis elimination (5), in the beginning of 2015, the Portuguese Ministry of Health implemented a national plan of universal access to HCV treatment (34, 35). In the framework of this programme, named Portal of Hepatitis C, 25309 patients have been authorized to initiate combinations of DAA-based treatments by the end of August of 2019 and 14023 patients have been cured, representing 96.5% of those who have completed the treatment (36).

By January 2021, the number of patients authorized for treatment had already increased to 28289, with 17385 patients having completed treatment with a success/cure rate of 96.6% again (37), which turns Portugal into one of the EU/EEA countries that met the European 2020 target of 90% of SVR among treated patients (Figure 1.6) (24).
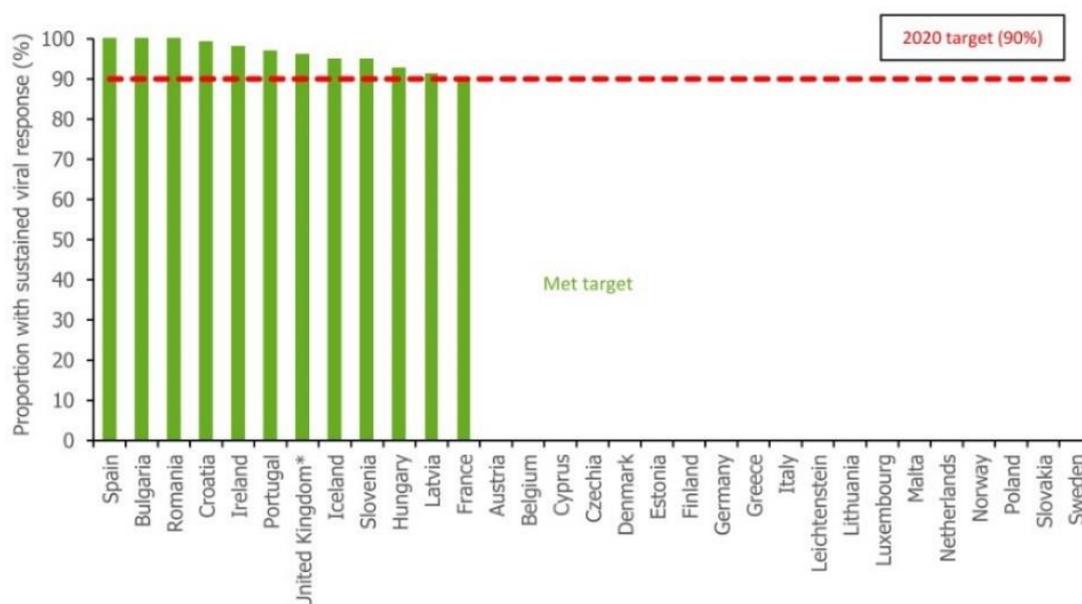
Figure 1.6 | Sustained virological response among HCV treated patients in EU/EEA countries, 2017. Graphic based on data from a 2019 report of ECDC (22); * represents data from England. (Image source: The sustainable development goals and hepatitis B and C in the EU/EEA, WHO 2021; licensed under CC BY 4.0).

However, patients who do not respond to DAA, although few, cannot be forgotten. It is important to understand what the causes of a poor or absent virological response are, that can be related to the host and to the viral strains. However, in Portugal, research about the molecular epidemiology of HCV strains has been scarce.

1.4.2. Hepatitis B

In Portugal there are important gaps in reporting hepatitis B data, especially about scale up of diagnostics and treatment. Nonetheless, some achievements have been done namely in implementing strategies to prevent vertical transmission and blood collection surveillance that are in line with goals of WHO for 2020.

1.4.2.1 Prevalence estimates in general Portuguese population

In 2018, the Portuguese General Directorate of Health – Direção Geral da Saúde (DGS), stated that only 174 cases of HBV infection were reported in the country, with higher frequency in people of the male gender and in age groups that were not covered by the vaccination, i.e., ages over 37 years (38). It also registered the occurrence of 16 new cases among blood donors, a reduction of 50% in relation to infected blood donors detected in 2016.

Portugal is considered a low endemicity country with a HBs antigen (HBsAg) prevalence in the general population that decreased along the years until achieving 0.02 – 1.45% in the decades of 1990 - 2014 (39-41). The last national serological survey from 2015/2016, estimated a chronic infection prevalence of 0,7% among men and 0.2% among women (29). Nonetheless, 2.9% of hospital admissions for liver cirrhosis in Portugal, between 2003 and 2012, were still related to hepatitis B (42).

1.4.2.2. Prevalence estimates in key Portuguese populations

When analyzing the situation by key populations, in 2018, DGS indicated the existence of 1.18% of patients on hemodialysis positive for HBsAg (38).

The DGS also gathered HBV data from SICAD regarding active users who consume illicit psychoactive substances, revealing that about 67% performed at least one HBsAg test in 2017 and 2018, with reactive results in 4.43% and 4.35% respectively. These values rise to 5.77% and 5.39%, respectively, when analyzing the population that injects drugs (38). Furthermore, the estimated prevalence of chronic hepatitis B infection among migrants from intermediate/high-endemicity countries in Portugal is around 9.0% (21).

1.4.2.3. HBV Birth dose Vaccination

Vaccination against HBV is the most effective way to prevent the disease, showing an efficacy of 95 - 99% (38). In Portugal, the vaccine has been part of the National Vaccination Program since 1994, initially aimed at young people between 11 and 13 years old and, since 2000, following the recommendation of the WHO, to all newborn children, with the scheme of administration of 3 doses: the first one after birth while still in the maternity hospital, the second at 2 months of age and the third at 6 months of age. In 2018, the hepatitis B vaccine reached 98% coverage at 12 months (38).

Portugal is one of the five EU/EEA countries that provide a universal birth dose of HBV vaccine. The most recent data on vaccine coverage of the birth dose, from 2017, revealed that together with Bulgaria, Lithuania, and Poland, Portugal have achieved the 2020 target of 90% birth vaccine coverage (Figure 1.7) (21).
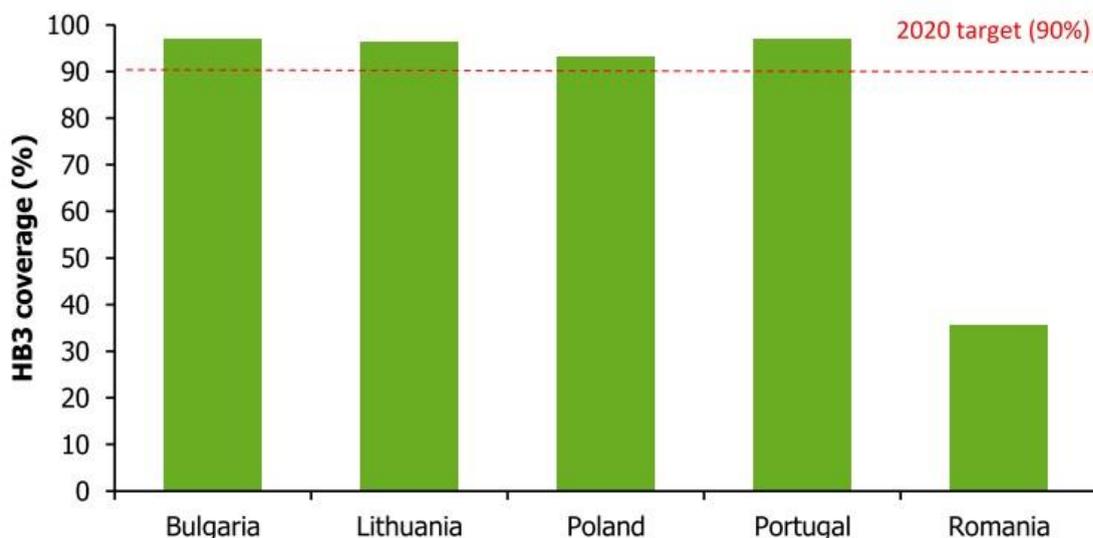
Figure 1.7 | HBV birth dose coverage in EU/EEA, 2017. Portugal is one of the four European countries that exceeded the WHO target for 2020. (Image source: Prevention of hepatitis B and C in the EU/EEA and the UK, ECDC 2021; licensed under CC BY 4.0).

1.5. Changes of infection patterns: examples from Portugal

In Portugal, hepatitis B has a higher prevalence in the migrant population (43). Individuals from endemic countries can bring rarer strains into Portugal. Together, the patterns of migration and HBV evolution rates may be influencing the epidemiology of HBV infection in the country

This also happens with HCV (43). Viral strains are usually grouped and classified into

genotypes (HBV) or subtypes (HCV) depending on their genetic similarity, and these groups can be further subdivided into subgenotypes and sub-subtypes, respectively. It is known that in some HCV cases, different subtypes respond differently to medicines that treat and cure the virus. In Portugal, it was observed in a study with 820 individuals under therapeutics with DAAs (ledipasvir/sofosbuvir or sofosbuvir) that 23 (2.8 %) did not respond, with most of these viral strains belonging to subtype 4 (34.8%), immediately followed by subtype 1 (30.4%) and subtype 3 (26.1%) (44). In this study, a higher unusual proportion of subtype 4 viral strains was found, which was thought to be related to immigration from African countries to Portugal. These strains modified the usual response to DAAs that commonly fail more frequently among 1a genotypes.

Knowing the distribution, molecular epidemiology and phylodynamics of these viruses, can help design informed strategies that can maximize the impact of interventions and ensure greater equity in their application. Currently, there are no studies describing the entry and dispersion of different HBV and HCV strains in Portugal, so it is not known which countries contribute to its spread in Portugal and whether there are specific times when this spread occurred more frequently.

Part II

The Infectious Agents of Hepatitis B and C

1.1. Viral Genomic Structure and Life Cycle

To understand such genetic diversity and issues related to HBV and HCV evolution, it is necessary to know the structure of the genome of these viruses and how they both replicate in the human body.

1.1.1 Hepatitis C virus

HCV belongs to the *Flaviviridae* family and *Hepacivirus* genus (45). It is a small enveloped virus that contains a single positive-stranded ribonucleic acid (RNA) genome with a nucleotide length of 9600 base pairs (bp) (46). The HCV genome contains a single long open reading frame (ORF) which encodes structural and non-structural genes and is flanked by 5' and 3' untranslated regions (UTRs) (Figure 1.8, A). (47, 48). The 5' UTR has a highly structured regulatory element called internal ribosomal entry site (IRES), that mediates the translation of a precursor protein of 3000 amino acids (aa) (Figure 1.8, B). This polypeptide is processed into 10 different proteins encompassing structural proteins (core, E1, E2 and p7) and non-structural proteins (NS2, NS3, NS4A, NS4B, NS5A, and NS5B), each one with specific functions in HCV life cycle (Figure 1.8 C).
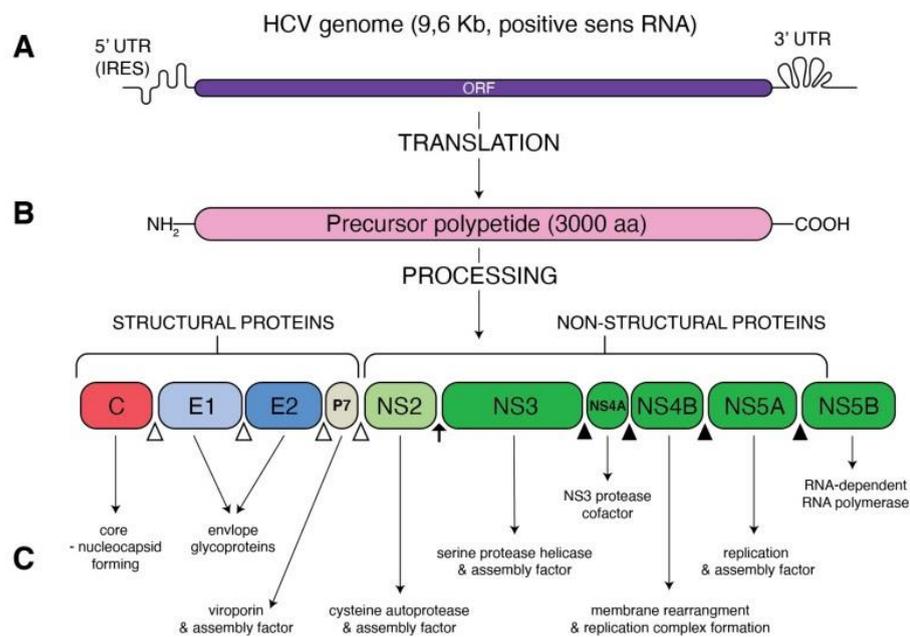
Figure 1.8 | Genome structure of HCV. HCV RNA (A) encodes a polyprotein (B) composed of about 3,000 aa. The core protein, the two envelope proteins and p7 are classified as structural protein, while NS2, NS3, NS4A, NS4B, NS5A, and NS5B are non-structural proteins. Their main roles are described in the figure (C) (Image from Sidorkiewicz *et al*, Metabolites, **2021;** Licensed under CC BY 4.0).

HCV viral particles circulate in the blood bound to lipoproteins, lipoviral particles (LVP), and due to different lipoprotein components of LVP circulating in the bloodstream, viral particles attach onto target cell membranes by sequential binding to lipoprotein receptors (Figure 1.9; attachment) (48). Then, the particles enter target cells, hepatocytes, via clathrin-mediated endocytosis (Figure 1.9; entry).
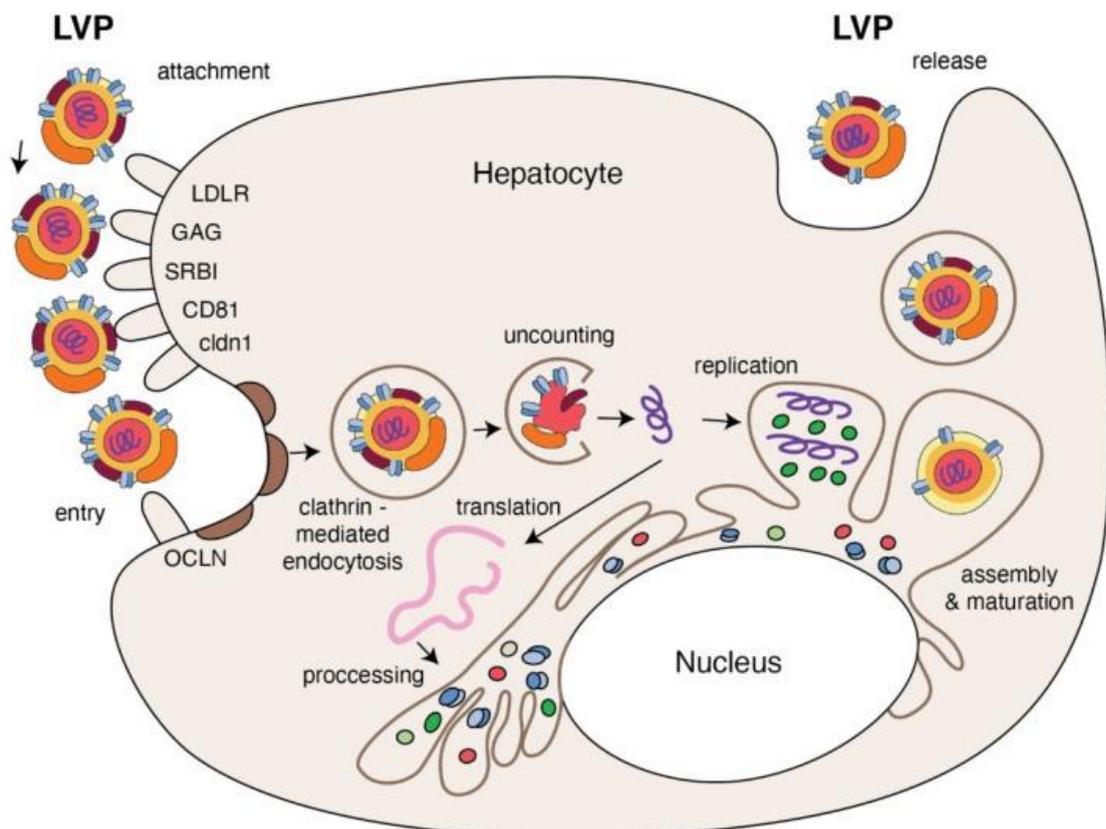
Figure 1.9 | Life cycle of HCV. Representation of HCV life cycle involving viral attachment, endocytosis, RNA uncoating, polyprotein translation and processing, RNA replication and, finally, viral assembly and release. ER, endoplasmic reticulum; GAG, glycoaminoglycans; LDLr, low density lipoprotein receptors; SRBI, scavenger receptors class B type I; CD81, tetraspanin CD81; cldn1, claudin-1 (tight junction proteins); OCLN, occludin.(Original image from Sidorkiewicz *et al*, Metabolites, **2021;** Licensed under CC BY 4.0).

During the uncoating event in the cell cytoplasm, HCV genome is released becoming available for translation of viral proteins via the 5'–IRES, that mediates binding of the HCV RNA to the host ribosomal subunits and cellular factors (Figure 1.9; uncoating and translation). The ORF is translated into a single 330 Kilodalton (kDa) viral precursor polyprotein at the rough endoplasmic reticulum (ER), which undergoes proteolytic

cleavages by viral and host proteases to generate the 10 viral proteins (Figure 1.9; processing).

The core nucleocapsid protein and the two envelope glycoproteins E1 and E2 form the N terminus of the polyprotein and are structural components of the HCV particles which are reported to function at viral entry into the cell (47). Non-structural proteins play important roles in viral replication and assembly (48, 49).

During HCV life cycle, nonstructural proteins and host factors form a multicomponent RNA replication complex within subdomains of the ER, called the membranous web (48), where positive-strand RNA is replicated by the virally encoded RNA-dependent RNA polymerase (Figure 1.9; Replication) via a negative strand intermediate. The last steps of this process involve the recruitment of the newly synthesized positive-strand viral RNA genomes to ER membranes, where genomic RNA is encapsulated into progeny virions, which are thought to bud into the ER lumen, thereby acquiring envelope proteins bound to the ER membrane before being released, via the Golgi apparatus, through exocytosis (50). Along this route, the virus matures from a high-density viroparticle to an infectious, low-density lipoviroparticle ready to infect naive target cells (48, 50).

1.1.2. Hepatitis B virus

HBV is a small enveloped deoxynucleic acid (DNA) virus, which belongs to the *Hepadnaviridae* family and *Orthohepadnavirus* genus (51). The infectious particle has an inner icosahedral capsid which contains a reverse transcribed, partially double-stranded circular DNA genome with approximately 3200 bp (51), being a partially double-stranded DNA molecule with a complete negative strand and a positive strand of variable length presenting a relaxed circular DNA (rcDNA) conformation (52, 53). It contains four overlapping and frame shifted ORFs, which are Pol (P), Core (C), Surface (S) and X, producing five viral RNA transcripts with different lengths, which are translated into 7 proteins (Figure 1.10) (52-54).

Pol is the longest ORF covering 70% of the viral genome and encodes the viral polymerase, which includes four functional sub-domains that are enzymes required for DNA synthesis: the polymerase, the reverse transcriptase (RT) RNase H, a spacer region, and a terminal domain (55, 56).
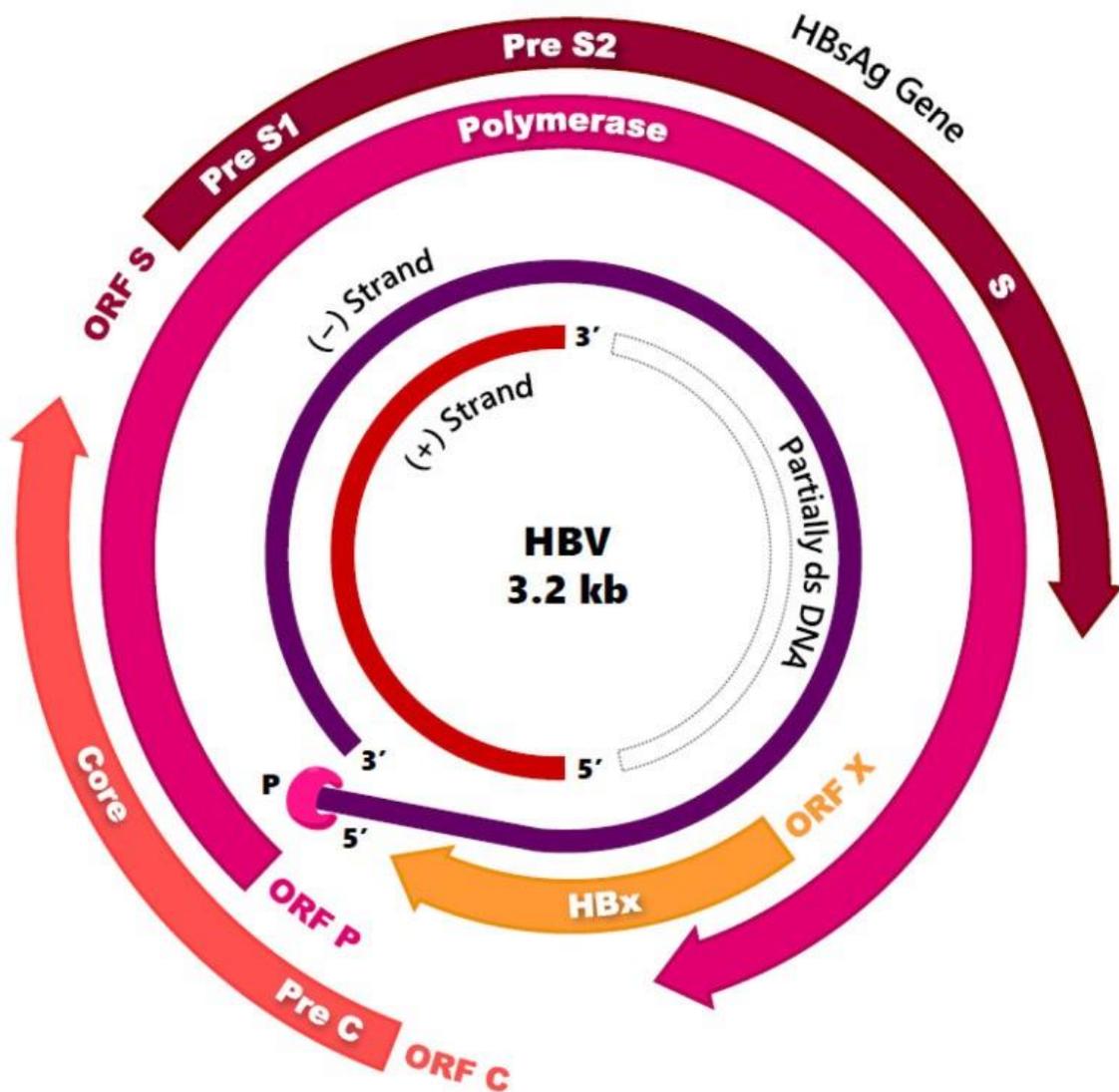
Figure 1.10 | Genome structure of HBV. Representation of the partially doubled-stranded circular genome of HBV with the four ORFs (colored arrows): P in dark pink, C in light pink, S in red and X in orange and the respective genes each one encodes annotated inside the arrows. (Original image from Prifti *et al*, Pharmaceuticals, **2021**; Licensed under CC BY 4.0).

The S-ORF encodes three HBV surface antigens: (HBsAg) proteins: large (LHB), medium (MHB), and small (SHB) proteins (56). The precore/core ORF encode for the core nucleocapsid protein, HBcAg, as well as the precore non-structurally antigen, HBeAg, necessary for the establishment of persistence and a marker of active infection (55). The X ORF encodes for HBx, the X protein, which is not a structural protein and is required for a covalently closed circular DNA (cccDNA) transcription being therefore a key regulator of HBV replication. It also confers a carcinogenic potential to HBV (57, 58).

When circulating in the body, the virus enters hepatocytes through host heparan sulphate proteoglycans (HSPGs) cell surface receptor binding (Figure 1.11).
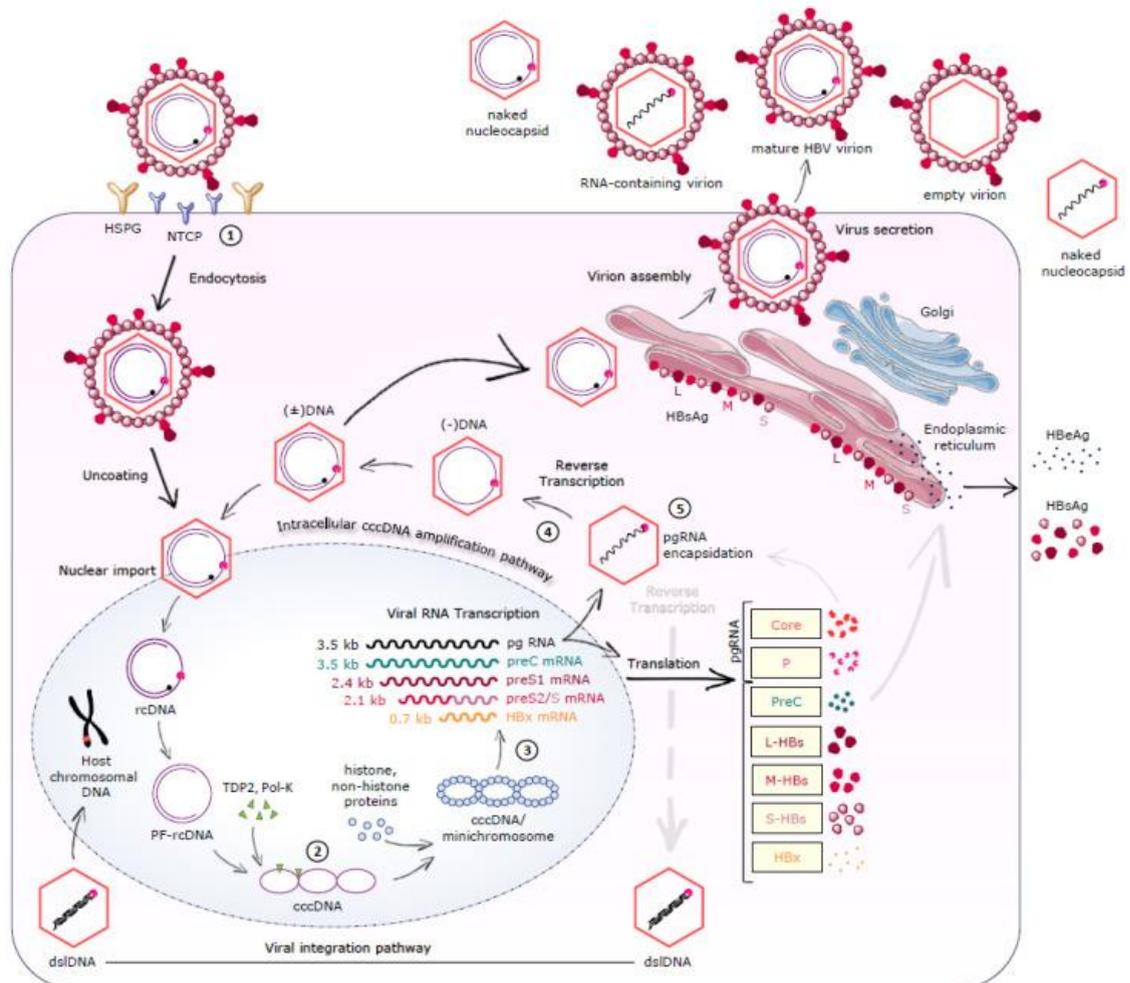
Figure 1.11 | Life cycle of HBV. Representation of HBV life cycle: Circulating viral particles attach to receptors in hepatocyte surface and enters the cell via endocytosis (n.º 1). Nucleocapsid enters the nucleus uncoating DNA, which is used to produce a cccDNA (n.º 2). This works as a template for transcription of a pgRNA and mRNA for viral proteins (n.º 3), which are translated in the cytoplasm. pgRNA is encapsulated with viral polymerase while starts the synthesis of the minus strand HBV DNA, followed by the plus strand (n.º 4, n.º 5). Finally, the nucleocapsid and surface proteins assemble in the ER and new viral particles are released to the exterior of the cell. NTCP; Na(+) taurocholate co-transporting polypeptide, HSPG; heparan sulphate proteoglycan, rc-DNA; relaxed circular DNA, PF-rcDNA; protein-free rcDNA, cccDNA; covalently closed circular DNA, pgRNA; pregenomic RNA, preC; precore, mRNA; messenger RNA, P; polymerase, L-HBs; large hepatitis B surface protein, M-HBs; middle hepatitis B surface protein, S-HBs; small hepatitis B surface protein, HBx; hepatitis B X protein, HBsAg; hepatitis B surface antigen, HBeAg; hepatitis B e antigen, dslDNA; double-stranded linear DNA (Original image from Prifti *et al*, Pharmaceuticals, **2021;** Licensed under CC BY 4.0).

31

Then it becomes internalized (endocytosis) due to high affinity interaction between HBV pre-S1 domain of L-HBsAg and Na(+)-taurocholate co-transporting polypeptide (NTCP) cell receptor (54, 55).

Viral surface proteins are removed (uncoating) and the nucleocapsid enters the nucleus releasing the rcDNA of HBV. Despite being a DNA virus, HBV replicates via reverse transcription using an intermediate RNA. In the nucleus, viral DNA is first converted into a complete double stranded genome and subsequently reconfigured as a cccDNA, a mini chromosome that serves as the template for transcription of HBV pre-genomic RNA (pgRNA) intermediate and messenger RNAs (mRNAs) for viral proteins. In addition to persisting as a minichromosome in the form of cccDNA, it also integrates into the host genome (53, 59, 60).

The mRNA molecules are transported into the cytoplasm, where they are translated into 7 viral proteins (small, medium, and large S proteins, core, e antigen, polymerase, and X protein). HBV RT produces a negative-strand DNA from pgRNA while it is being coated by the viral nucleocapsid. The pgRNA template is degraded by RNase H, and then synthesis of the positive-strand DNA is initiated. HBV DNA is repackaged in the relaxed form and new virions assemble with viral surface proteins in the ER before being released from the cell into the blood (52).

1.2. Diversity of HCV and HBV

As discussed above, the replication of HBV and HCV is dependent on viral polymerases that act in an important reverse transcription step from both life cycles, neither of them can correct the errors they introduce during the synthesis of new viral sequences (61-63). This fact leads to sequence heterogeneity (62) and due to this nucleotide genome divergence among the viral population, these viruses can be classified into different groups, called genotypes, depending on their genetic distance. HBV strains are grouped into 3 levels: a) genotypes; b) subgenotypes and c) clades (64-66). Major genotypes with a nucleotide divergence above 7,5% (67, 68) are further subdivided in more than 30 sub-genotypes with a divergence between 4% and 7,5% (67-69). Clades have less than 4% of nucleotide divergence (65). Despite of having less than 4% of nucleotide divergence over the full-length genome, some HBV strains does not assign in any independent subgenotype, and they are called "quasi-subgenotype" (64). They cannot be assumed as

clades because clades are subclasses of defined subgenotypes, and these strains do not clearly belong to a subgenotype. Furthermore, the nucleotide divergence of "quasi-subgenotype" is almost more than 3% and less than 4%, whereas nucleotide divergence between clades and other clusters in their main branch is less than 3% (64).

On the other hand, HCV strains are also classified in 3 levels, called genotypes, subtypes and clades. The genomes of the HCV isolates that differ by about 30% at the nucleotide level are classified into different major genotypes, which are also subdivided into subtypes when genomes differ between 15% and 30% (70-72). Clades were only identified in subtype 1a (73).

As of May 2019, HCV is classified into 8 genotypes and 90 different subtypes (Figure 1.12) (72, 74), where 1a, 1b, 2a and 3a are the most widely distributed subtypes across the world (75).

Naturally occurring intra or intergenotypic recombinants are rare, but a particular 1b/2k intergenotypic recombinant has spread to become of epidemiological importance (76).
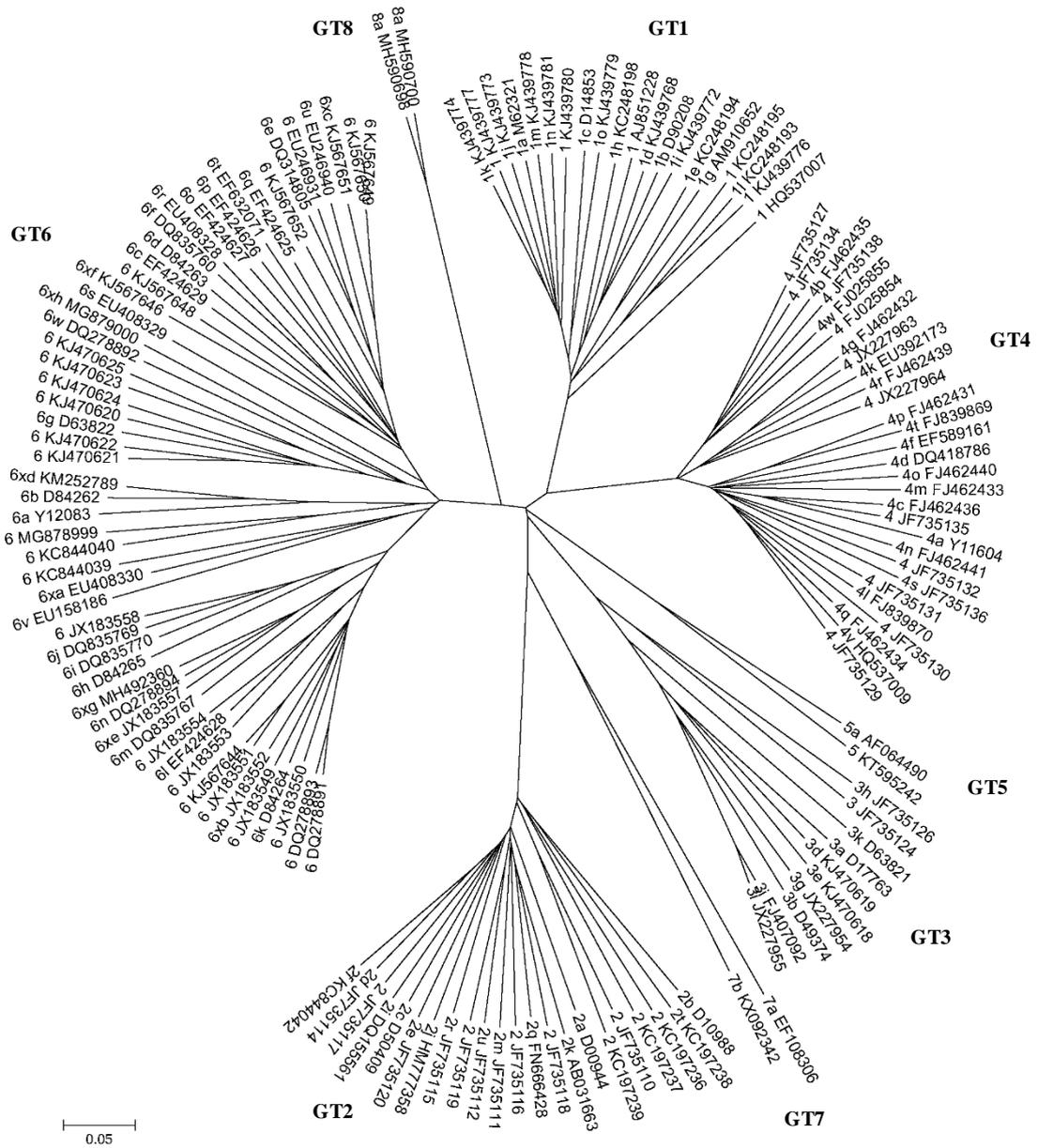
Figure 1.12 | Neighbor-joining phylogenetic tree showing the confirmed HCV genotypes and subtypes as proposed by International Committee on Taxonomy of Viruses (ICTV). Clades corresponding to each genotype were supported by 100% of bootstrap replicates. GT; genotype (Original Image from Smith *et al*, ICTV website, 2019; licensed under CC BY-SA 4.0)

HBV major genotypes are classified from A to I (68), and a 10th putative genotype J (77) (Figure 1.13). There are three A subgenotypes A1, A2, A4, and a quasisubgenotype A3, four B subgenotypes B1, B2, B4, B5, and a quasisubgenotype B3, fifteen C subgenotypes C1, C3 - C16 and a quasisubgenotype C2, six D subgenotypes D1 – D6, four F subgenotypes F1 - F4 (78) and two I subgenotypes I1 and I2 (79).
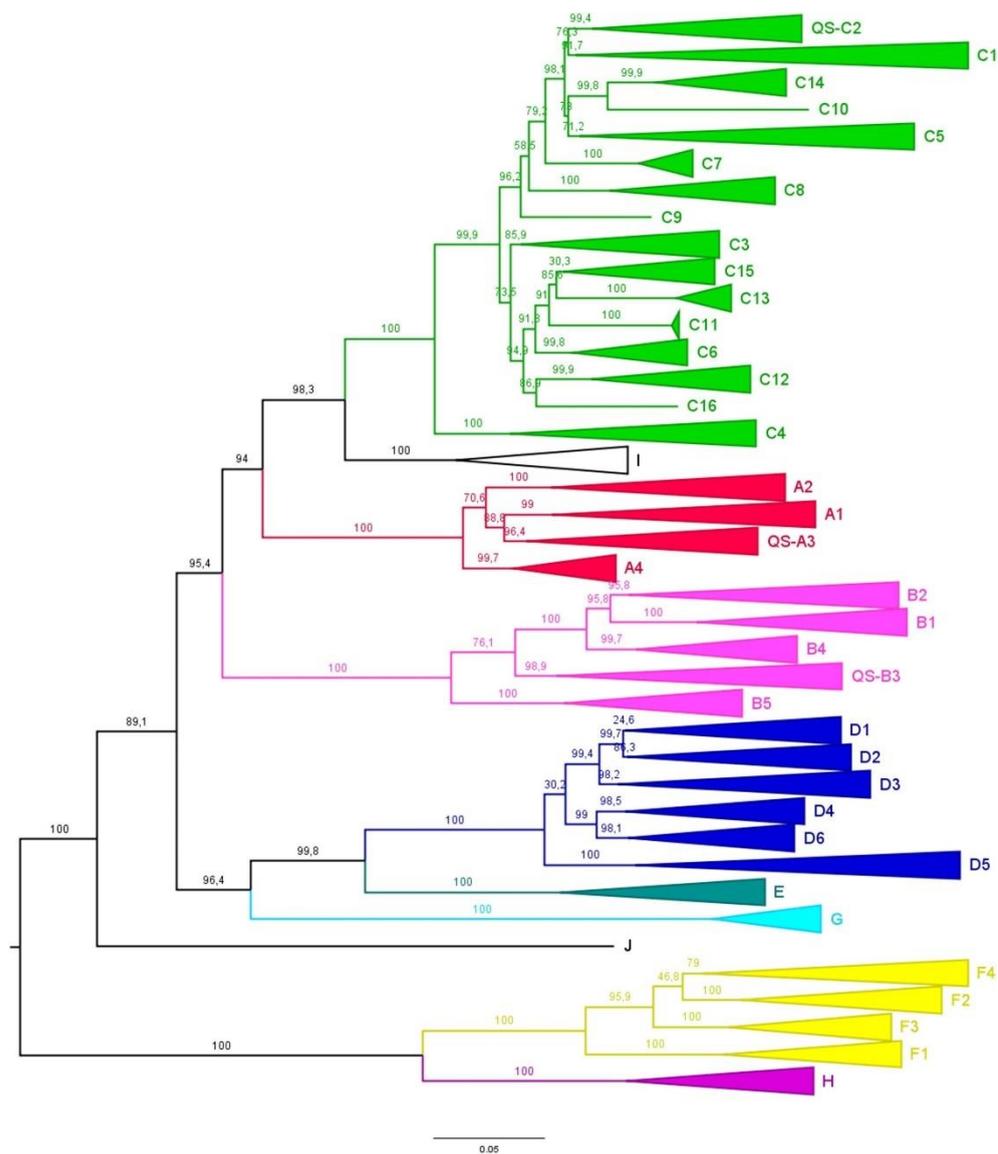
Figure 1.13 | Phylogenetic analysis of HBV complete genome sequences representative of genotypes A – J. Maximum likelihood phylogenetic tree of complete HBV genome sequences. The branches are colored according to the genotypes. Subgenotypes are indicated next to their respective clusters. The numbers in branches indicate the statistical support (aLRT value). (Original Image from Araujo *et al*, Frontiers in Microbiology, 2020; licensed under CC BY)

The potential 10th genotype, genotype J, was found to be a recombinant of genotype C and gibbon HBV in the S region and that may represent an independent cross-species transmission (77).

35

1.3. Worldwide distribution of HCV and HBV genotypes

1.3.1. Hepatitis C virus

HCV genotypes 1 and 3 are the most prevalent worldwide, followed by genotypes 2, 4, 6 and 5 in this order (75). Genotypes 7 and 8 were more recently found and were described only in a few countries (72, 80, 81).

Globally, genotype 1 is estimated to account for more HCV cases than any other genotype, with over one-third of the cases located in East Asia (75). HCV genotype 3 is the second most common and accounts for approximately three-quarters of the cases in South Asia. However, it also dominates parts of Scandinavia (Figure 1.14).
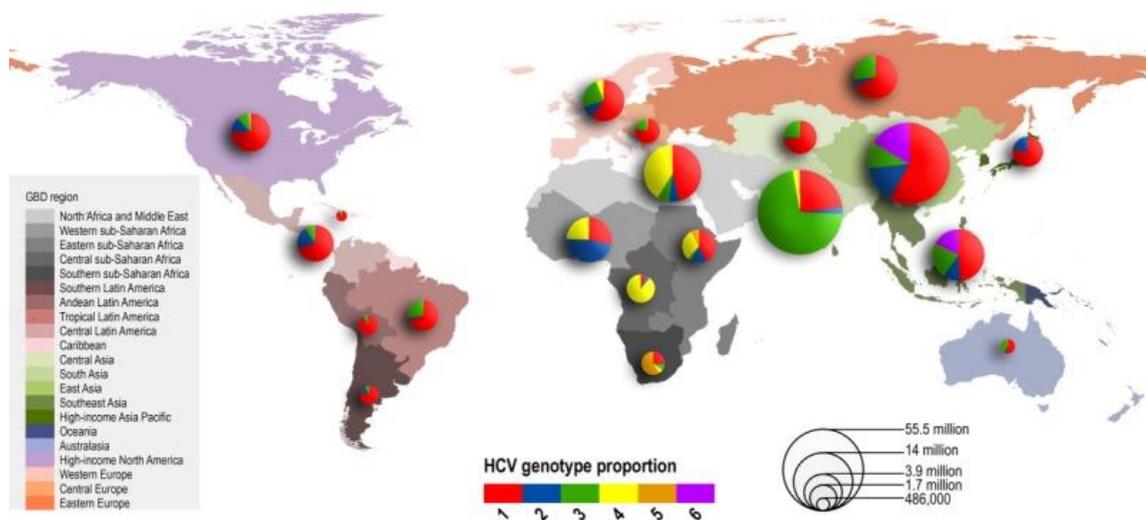
Figure 1.14 | HCV worldwide proportion by Global Burden of Disease (GBD) region. This image results from the combination of overall HCV prevalence estimates from the GBD project with data from studies published between 1989 and 2013 reporting HCV genotypes. Size of pie charts is proportional to the number of seroprevalent cases as represented by the circles in the bottom right. (Image from Messina *et al*, Hepatology **2015**; licensed under CC BY).

Genotypes 2, 4, and 6 are responsible for most of the remaining cases of HCV worldwide, with genotype 2 prevailing in East Asia and West Africa, genotype 4 showing the largest number of cases in North and Central Africa and in the Middle East, while genotype 6 is more frequent in East and Southeast Asia (75). Genotype 5 is responsible for less than 1% of HCV cases globally with the great majority occurring in Southern and Eastern sub-Saharan Africa (75).

Several HCV subtypes, especially 1a, 1b and 3a and, to a lesser extent, 2a, 2b and 2c, are globally distributed and for this reason are considered as pandemic (82). Many other HCV

subtypes are considered "endemic" strains; these are comparatively rare and have circulated for long periods of time in more restricted regions (75). In Egypt, the dominant subtype is 4a which seems to have resulted from its spread due to unsafe injections during past anti-schistosome public health campaigns (83). Other subtypes have been amplified locally and regionally after being introduced into networks of injection drug users, for example, subtype 4d in Europe (84, 85) and subtype 6a in Vietnam and Hong Kong (86). Subtype 5a appears to have originated in South Africa but has a second region of occurrence in Belgium (87). Genotype 7 seems to be restricted to a few cases in Democratic Congo Republic (80, 81) and genotype 8 was identified only in four Indian individuals resident in Canada (72).

1.3.2. Hepatitis B virus

HBV genotypes have distinct geographical distribution showing similar patterns between countries of the same world region but strongly varying between different parts of the world (88).

Genotypes A and D are the most frequent in Europe with the former prevailing in Northwestern Europe and sub-Saharan Africa (Figure 1.15) (88, 89). In this last region, the subgenotype A1 is the most frequent of all A subgenotypes (90).
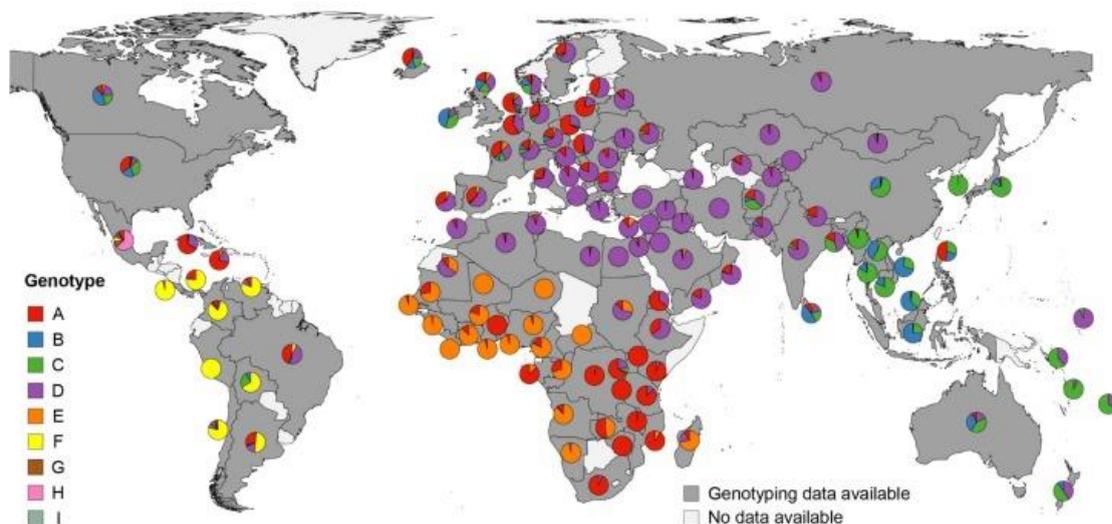
Figure 1.15 | HBV worldwide distribution. The proportional HBV genotype distributions in the respective countries are indicated in pie charts (Image from Velkov *et al*, Genes **2018**; licensed under CC BY 4.0).

Genotype D is also common in Northern Africa and in some parts of Asia (88), such as India and Central Asia (67, 88, 89). The subgenotype D3 is widespread globally, but the subgenotype D4 has also been described, namely in the region of Latin America (Venezuela, Mexico, and Brazil), the Caribbean (Cuba, Haiti, Martinique) and Oceania (90, 91). Genotype C is the most frequent genotype in Eastern and Southeastern Asia, as well as Australasia and Oceania, but genotype B also represent a high proportion in China, Southeast Asia, and Australia (67, 88, 89) with subgenotype B5 (formerly B6) prevailing in a Canadian Inuit population (92, 93). Although less frequently, subgenotypes D3, D4, and A2 were also found in these arctic populations in this order of prevalence (92). The Philippines are an exception in Asia region with about half of infections caused by genotype A (88). In Sub-Saharan Africa, genotype E prevails in its western and central regions (88, 94). It shows decreasing proportions towards Eastern Africa, where, except for Madagascar that also is dominated by genotype E, mainly genotype A is found (88). Genotypes F, G and H are found in Latin America, but rare in other parts of the world. However, Brazil and Mexico differed from the other countries in the region because genotypes A and D dominate in Brazil, and uniquely in Mexico, genotypes G (10.2%) and H (63.3%) are found in relevant numbers (67, 88). However, genotype H is already found in North America and Asia (66) whereas HBV genotype F is also found in Central America, as well as in Alaska (67). Genotype G is reported in France and United States where it was discovered (95), but also in some other European countries (96) such as Belgium (97), Netherlands (98), or Germany (99) and in Central America as well as in Asia and Africa (96). Genotype I has been reported in Vietnam and Laos (79). The newest HBV genotype J has been identified in the Ryukyu Islands in Japan (77).

Contrastingly to Latin America, Northern America shows a predominance of genotypes A, B, C and D, whereas in the Caribbean, mainly genotype A, and, to a lesser extent, genotype D are found (Figure 1.15).

1.4. Genotype and subgenotypes/subtypes distribution in Portugal

Limited information exists regarding HCV and HBV genetic diversity in Portugal. Nonetheless, some information about HCV genotypes and subtypes has been published in studies performed between 1998 and 2014 (100-104). These studies have been carried out in different patient populations, such as individuals who were followed up in external

hospital consultations (101, 102, 105), individuals who carried out routine testing in laboratories (104) and populations of PWID (103). The percentages of each genotype vary according to the population, but it is a consensus for all studies that most HCV infections in Portugal are caused by genotype 1 (HCV/1) followed by genotype 3 (HCV/3), genotype 4 (HCV/4) and genotype 2 (HCV/2). However, it seems that until the year 2000 subtype 1b (HCV/1b) was the most prevalent (100, 101), but it was gradually overtaken by subtype 1a (HCV/1a) which is now considered as the most prevalent in the country in the most recent published studies (103, 104). Prevalence estimates vary between 45 to 65% for HCV/1, 27 to 72% for HCV/3, 9% to 25% for HCV/4, and around 2,5% for HCV/2. Except for HCV/2, which was absent in PWID, this key population was the one with the highest prevalence rates (102-104).

As for HBV, a previous review of abstract proceedings from national meetings, reports and peer reviewed articles published between 2004 and 2014 tried to collect dispersed information on genotypes distribution along the country (39). The gathered studies were performed in a great diversity of populations such as blood donors, pregnant women, PWID, prisoners, children and adults attending hospitals, individuals who carried out routine tests in laboratories, and health care professionals (39). Independently of the populations studied, HBV genotype D (HBV/D) was consistently identified as the most prevalent genotype in chronically infected Portuguese patients, followed by HBV genotype A (HBV/A) (39). This also revealed that HBV/D was more common in the North of Portugal (59.5 – 73%) than in Central-Southern (24 - 54.4%) whereas HBV E genotype (HBV/E) was more common in Central-Southern (10 - 62%) than in Northern Portugal (1 - 4%) (39, 105). This aspect probably reflects different migration patterns to each region of Portugal (39). HBV Genotypes C and F (HBV/C and HBV/F) have also been reported in the North of the country but are considered rare cases (105). Previously unpublished data of a small patient population of Hospital Santa Maria in Lisbon also indicated the circulation of minority HBV genotypes B (HBV/B), C and F (106).

This was the unique study in Portugal which identified HBV subgenotypes D4 [(HBV/D4), (25,4%)], D3 [(HBV/D3), (19,3%)], A1 [(HBV/A1), (17,5%)] and HBV/Æ (17,5%) as the most prevalent in this population and a lower prevalence of HBV subgenotypes F2 [(HBV/F2), (2,6%)], A2 [(HBV/A2), (1,8%)], D1 [(HBV/D1), (1,8%)], D2 [(HBV/D2), (0,9%)], A3 [(HBV/A3), (0,9%)], B2 [(HBV/B2), (0,9%)] and C1

[(HBV/C1), (0,9%)] (106).

In Portugal, it is becoming evident along the years that hepatitis B disproportionally affects the migrant population, and the patterns of migration may be influencing the epidemiology of HBV infection in the country (35, 39).

Historical migration patterns to and from Portugal may have shaped the molecular epidemiological profile of HBV and HCV infection in the country (35, 39) and, therefore, it is important to investigate the epidemiological history, transmission dynamics and origin of the strains circulating in Portugal, as well as the role of this country in the global spread of these viruses. However, until now, this type of study has never been carried out in Portugal.

Part III

Viral Evolution

1.1. Why trace viral evolution?

During the last 40 years there have been major advances in the study of virus evolution (107). In a relatively recent past, genetic distances and phylogenetic trees inferred using various evolutionary models, were regularly used to assign the viral genotypes but ended up proving its usefulness by going further and allowing to define circulating recombinant forms (CRFs) (108), to discover mosaics and complex form of viruses (109, 110). In the last years, several new methods for inferring phylogenetic trees have been introduced that allow to study the ancestral relationships of viruses sampled from a population. These methods are based on Bayesian theory introducing the statistical phylogenetic framework with two important concepts: the molecular clock hypothesis and the coalescent theory (111).

Phylogenetics are now a powerful field to analyze at the genomic scale the emergence of viral diversity, turning it into a key tool in the response to infectious diseases outbreaks, as demonstrated in the recent epidemics of Middle East Respiratory Syndrome Coronavirus (MERS-CoV) (112), Ebola (113), Zika (114), various forms of influenza virus (115-117) and in the present days with Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV2) (118). Today's phylogenetic approaches in virology can help revealing more than viral genotypes, shedding light into the patterns, processes, and rates of cross-species transmission (i.e., spillover events) or intra-species dissemination, as well as its determinants (119, 120). As an example, molecular phylogenetics elucidated about SARS-CoV2 mechanisms of viral emergence and spread, contributing to provide clues about the possible paths of virus transmission from bats to humans and offering some clues about viral origins, which will be important to contain future epidemics (118). In addition, through phylogenetic tools it is now possible to incorporate epidemiological data with molecular sequences into a single statistical framework creating an integrative analysis that allows to characterize the transmission dynamics of fast evolving viruses, a field called phylodynamics. This approach allowed for scientific advances in the determination of temporal and geographical origins, evolutionary history, and ecological risk factors associated with the growth and spread of viruses (121). Although viral spread is often attributed to human mobility, factors such as population growth related to urbanization and accessibility can also play an important role, as has been demonstrated for the emergence of human immunodeficiency virus type 1 (HIV-1) group M (122), and

subtypes A and D in East Africa (123) and CRF02_AG in regions of the Congo River Basin (124).

In rapidly evolving pathogens, such as HIV, HCV and even HBV, sequence evolution occurs simultaneously with geographic dispersal and this interaction characterizes a spatial phylodynamic process that can be recovered from genomic data using phylogeographic analysis. The importance of phylogeography lies not only in the historical perspective of the viral evolution, but also in its potential to predict patterns of emergence and dissemination of infectious diseases. Furthermore, it can help to locate the habitat of key reservoir species and geographic areas from which new infections are likely to emerge and spread, as well as – when combined with sampling from animal hosts - elucidate the impact of animal movement and/or human mobility on viral disease spread (125).

1.1.1 Phylogenetics

In phylogenetics, evolutionary relationships and processes of evolution are established and studied through the analysis of sequence data. When applied to virology, molecular phylogenetics have been useful to study the origins, emergence, and spread of fast evolving viruses (126), since their genetic variation accumulates within an epidemiological time frame. This adds power to phylogenetic analysis allowing phylodynamic inferences to be made with great precision (111). Most of the viruses with high intrinsic mutation rates are RNA viruses whose replication happens either through a RNA-dependent RNA polymerase or a reverse transcriptase (retroviruses), both of which lack capacity for proofreading or error correction (127). Such frequent mutation rates, combined with rapid rates of replication, provides RNA viruses with high overall rates of evolutionary change, which may vary depending on the group to which they belong according to the Baltimore classification (Figure 1.16) (127). Single stranded RNA (ssRNA) viruses, such as HCV, are those who present the higher rates of nucleotide substitution and thus a faster evolution, varying between 10^{-6} and 10^{-4} substitutions per nucleotide per cell infection (s/n/c) (127, 128). HCV has an average mutation rate around 3.8×10^{-5} s/n/c (127).

Baltimore classification also includes DNA viruses. Most of them have lower rates of variation, between 10^{-8} and 10^{-7}, meaning that evolutionary change cannot normally be observed in real time (127). Double stranded DNA (dsDNA) viruses, which replicate

using DNA polymerases with added repair mechanisms, have nucleotide substitution rates among 10^{-7} - 10^{-9} s/n/c, and thus evolve much more slowly than their RNA counterparts (127).
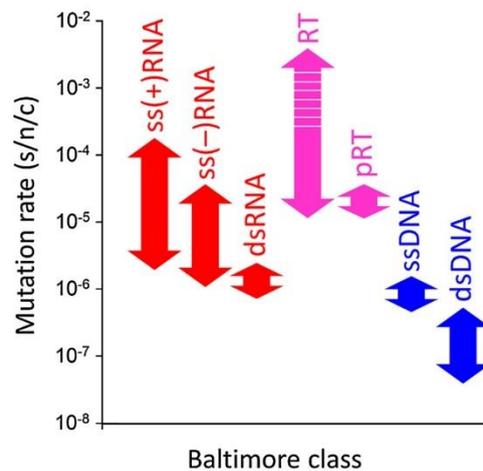
Figure 1.16 | Range of mutation rate variation across viruses. The seven groups of viruses are represented according to Baltimore classification (ss single-strand, ds double-strand, RT retroviruses, pRT para-retroviruses). In the RT group, all mutation rates fall in the nonhatched arrow region except the HIV-1 mutation rate measured in cellular DNA, which order of magnitude is higher than the rate measured in plasma. (Original Image from Sanjuán *et al*, Cellular and Molecular Life Sciences **2016** (only part A of the image was used); licensed under CC BY 4.0).

HBV which has a dsDNA but uses an RNA intermediate molecule in its replication cycle, belongs to a special group of viruses, the para-retrovirus group according to Baltimore classification (Figure 1.16). These viruses have specific nucleotide substitution rates that are superior to those of other dsDNA viruses because are complicated by other factors. In the case of HBV, the most notable of these factors is the small size and overlapping nature of its genome. According to this scale its rate is around 10^{-5} s/n/c (127).

Most estimates of the rate of evolutionary change in HBV in published literature have been close to 10^{-5} substitutions per site per year (s/s/y), but a recent study re-estimating the viral evolutionary rate in known transmission chains, observed rates in the order of 10^{-3} to 10^{-4} s/s/y (129).

1.1.1.1 Representation and interpretation of phylogenetic trees

Assuming there is a shared common ancestry, viral molecular evolution can be visualized through the reconstruction of phylogenetic trees, where a descendant-ancestor relationship between sequences can be represented through internal and external nodes and branches. The external nodes of a tree, usually called tips, represent viral sequences under study, while the internal nodes represent unobserved ancestral sequences (Figure 1.17).).
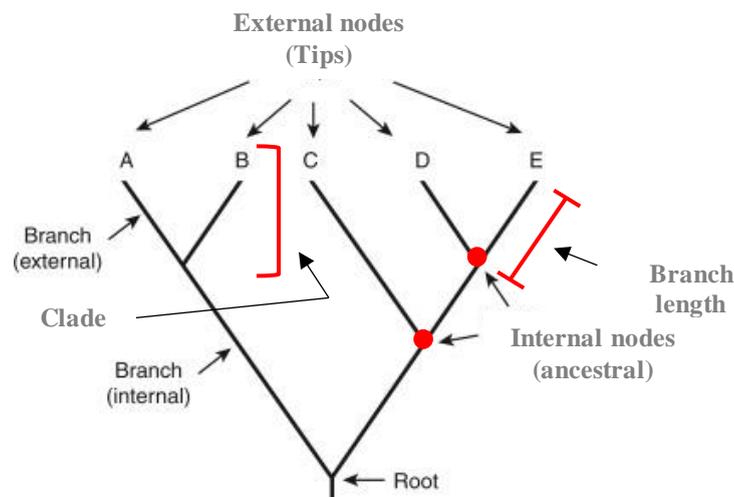
Figure 1.17 | Components of a rooted phylogenetic tree. Letters A to E are sequences under evolutionary study.

This aggregate of ancestors and descendants forms groups which are called clades. Depending on the process of tree reconstruction and on the chosen models, branch lengths represent the amount of evolution occurring between nodes that can be expressed in units of nucleotide substitution per site in the genome or in units of time (years, months, or days. The presence of a root in the phylogenetic tree represents the common ancestor of all the sequences in the tree and the branches formation indicates the direction of evolutionary history starting from the root (common ancestor) to the tips (studied sequences).

1.1.1.2 Reconstructing phylogenies – traditional vs Bayesian approaches

When studying viral evolutionary relationships, following the collection of sequences, the next step consists of downloading other relevant sequences from nucleotide sequence

databases and build an alignment of homologous sequence sites through multiple sequence alignment (MSA) tools. A correct MSA guarantees that nucleotides in each column of the alignment are related to each other by positional homology. The next step consists of choosing a model of sequence evolution, which depends on the objectives of the analysis and the complexity of the data. Different tree reconstruction methods can include for example neighbor-joining (NJ) or maximum-likelihood (ML) reconstruction – which are faster - or Bayesian approaches using Markov Chain Monte Carlo (MCMC). While NJ and ML reconstruct the tree that was considered the most correct, that can later be tested for reliability/confidence in each clade using bootstrapping processes; Bayesian approaches give us a sample of trees with the higher posterior probability, that can later be summarized in one Maximum-Clade Credibility Tree (MCCT) (Figure 1.18) (130).
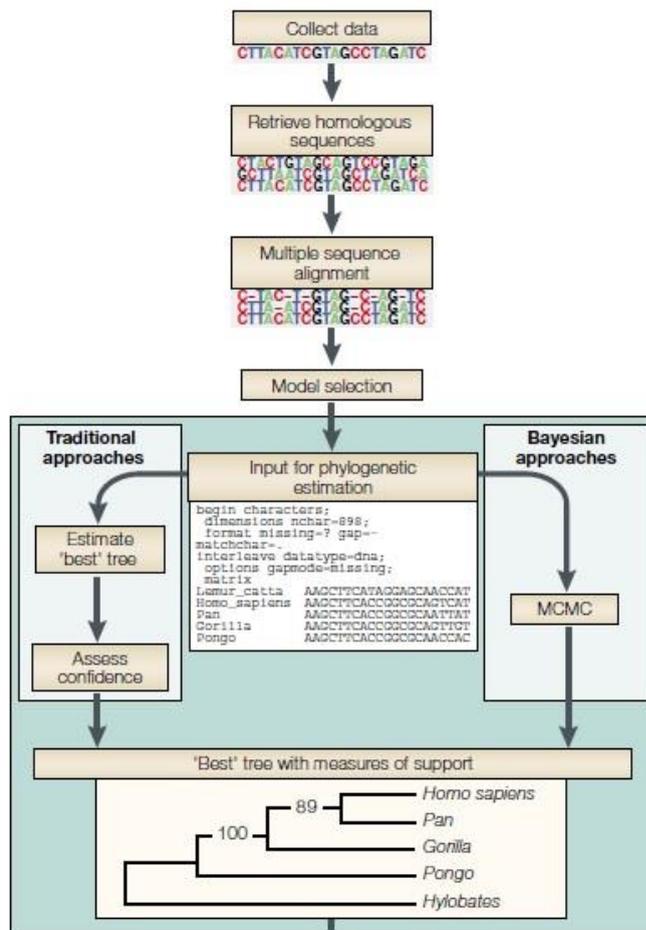
Figure 1.18 | The phylogenetic inference process. After alignment of collected data with retrieved homologous sequences, a phylogenetic tree can be inferred from data using either traditional or Bayesian approaches (Original Image from Holder and Lewis, Nature Reviews Genetics 2003; Licensed under © 2003 Nature Publishing Group, but reprinted by permission from Springer Nature under the license number 5124700873900 from 9 Aug 2021).

The most common traditional approaches to reconstructing phylogenies are either the neighbor-joining (NJ) approach or tree searches that use an optimality criterion such as parsimony or maximum likelihood (ML) (130, 131). The NJ converts the DNA sequences into a pairwise distance matrix that represents an estimate of the number of changes that have occurred along the branches between each pair of sequences. NJ is an extremely fast method to estimate of phylogenetic trees that does relatively well on sequences that have diverged recently. However, when the goal is to infer older relationships, it can perform worse than other methods. When situations as back mutations occur, such as nucleotide changes like A – C – A, or complex pathways such as C – G – C – A, these are unseen

events where the observed differences between sequences are not accurate reflections of the evolutionary distances between them. These homoplasies - multiple substitutions or back mutations - at the same site obscure the true distance and make sequences seem artificially close to each other (130).

In contrast to the distance matrix process, the parsimony method compares all sequences at the same time by calculating a score for each possible generated tree (131). This type of sequence comparison also occurs in ML and Bayesian methods. However, in the parsimony method in particular, the score is simply the minimum number of mutations that could possibly explain the differences in the sequence data (130). Thus, parsimony refers to choosing between trees based on which one requires the fewest possible mutations to explain the data (130, 131). However, it has been shown that parsimony is not necessarily a realistic evolutionary assumption. Given that the mutational pathway along the evolutionary tree is unknown to us, we should consider all the possible paths that could explain the data and therefore ML consider all possible mutational mappings (130).

Accurately reconstructing the relationships between sequences that have been separated for a long time, or are evolving rapidly, requires a method that corrects for multiple mutational events at the same site (homoplasies or saturation). ML provides such a method, by analyzing how well a hypothesis predicts the observed data and choosing the tree that has the highest probability of producing the observed sequences (this probability corresponds to the Likelihood of the tree) (130).

Bayesian approaches are the most recent in phylogenetics. Its primary analysis produces a sample of trees with the higher posterior probability, as well as measures of uncertainty for each clade of the tree (130). Before the analyses of the data, parameters are assigned a prior distribution (the probability an event will happen before you taken any new evidence into account), which is combined with the data (likelihood) to generate the posterior distribution (the probability an event will happen after all evidence or background information has been considered). All inferences concerning the parameters are then based on the posterior distribution. In the past two decades Bayesian inference (BI) has gained popularity thanks to advances in computational methods, especially MCMC algorithms (131).

To describe the data, NJ, ML, and BI all make use of a nucleotide substitution model and are therefore model-based, whereas parsimony does not have an explicit model and its assumptions are implicit (130).

1.1.1.3 Nucleotide Substitution Models

Pairwise sequence distances are calculated assuming a Markov Chain model of nucleotide substitution. Some commonly used models are illustrated in figure 1.19. The JC69 model assumes equal base frequencies and an equal rate of substitution between any two nucleotides whereas the K80 model assumes equal base frequencies but different rates for transitions and transversions. The assumption of equal base frequencies is relaxed in the HKY85 model and general-time-reversible (GTR) model both assuming variable base frequencies. In addition, GTR model allows all six exchangeability rates to differ (111).
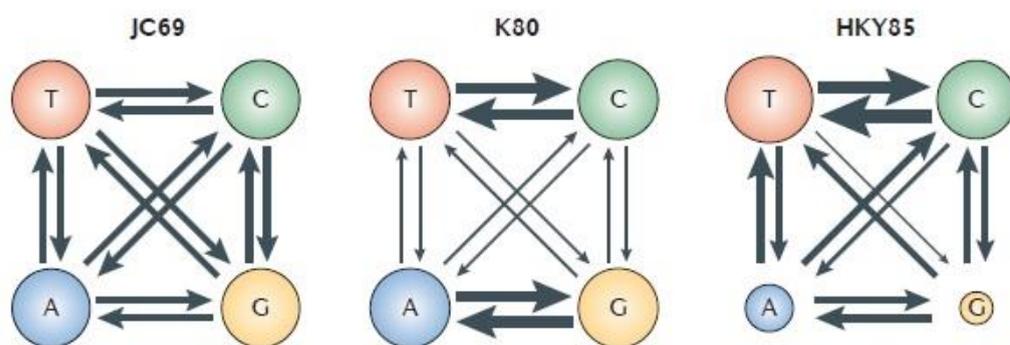
Figure 1.19 | Markov models of nucleotide substitution. The thickness of the lines indicates the substitution rates of the four nucleotides (T, C, A, G), and the sizes of the circles represent the nucleotide frequencies when the substitution process is in equilibrium. Note that both JC69 and K80 predict equal proportions of the four nucleotides (Original Image from Ziheng Yang and Bruce Rannala, Nature Reviews Genetics, 2012; Licensed under © 2012 Nature Publishing Group, but reprinted by permission from Springer Nature under the license number 5124701242051 from 9 Aug 2021)

Depending on the goal of a study, for instance when a phylodynamics or a phylogeographic analysis is the aim, the process of choosing an evolutionary model consists of a combination of a specific reconstruction approach with a nucleotide substitution model, as well as on choosing a molecular clock model, a coalescent model and/or a spatial diffusion model (125, 132).

1.1.1.4 Molecular clocks and coalescence in viral dynamics reconstruction

A molecular clock is a statistical model that describes the relationship between time and the genetic distances among nucleotide sequences. In contrast to older molecular clock models that required the assumption that the rates of nucleotide change were constant through time (strict molecular clocks), contemporary models assume evolutionary rates can vary as time goes by along the branches of the tree (relaxed molecular clocks) (132). A simple phylogeny has no timescale, so the branch lengths only represent the genetic divergence from the ancestor (Figure 1.20, a). However, when viral sequences have been sampled at known time points and if this phylogeny is calibrated using a molecular clock model, then the branches lengths will represent evolutionary rate calibrated with units of time (ex: nucleotide substitution/site/years), allowing to estimate the timing of branching events, including that of the common ancestor (Figure 1.20, b).

Furthermore, when incorporating space diffusion models into this MCMC framework, it is possible to reconstruct the phylogeography of fast evolving viruses. Combined spatial and genetic analyses not only reveals the location of origin of emerging infections but can also discern the route of transmission and the rates of geographic spread. In addition, combining spatial and temporal information provides details of when the dispersions occurred from one point to another (viral diffusions calibrated across time and space) (Figure 1.20, c).
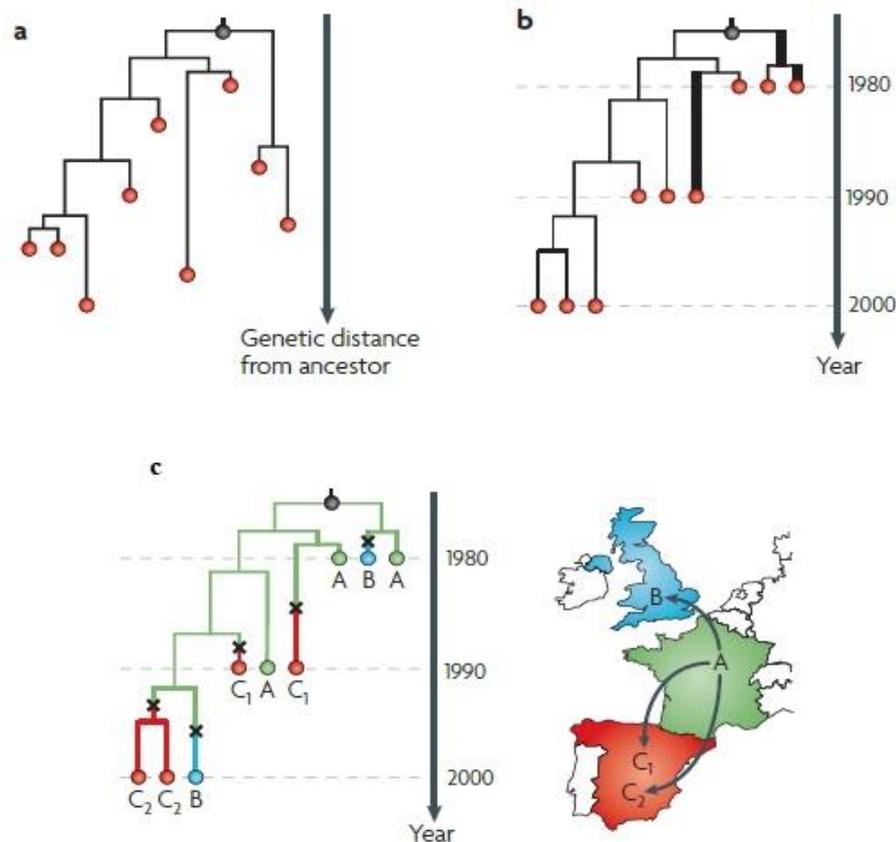
Figure 1.20 | Processes of phylodynamic analysis. a) Simple phylogenetic tree. Branches lengths represent genetic distance between samples and their ancestors. **b)** Phylogenetic tree combined with a molecular clock. Branches lengths represent units of time (years). **c)** Phylogenetic tree combining temporal and spatial information of samples. Branches colors represent different locations. The nine sequences were sampled from France (green, A), the United Kingdom (blue, B) and two locations in Spain (red, C1 and C2). Statistical methods can be used to reconstruct the history of viral spread, so that each branch is labelled with its estimated geographic position. Combining the spatial and temporal information provides further insights — this hypothetical virus spread from France to location C1 years before independently arriving at location C2. (Original image from Oliver Pybus and Andrew Rambaut, Nature Reviews Genetics 2009; Licensed under © 2009 Macmillan Publishers Limited. All rights reserved, but reprinted by permission from Springer Nature under the license number 5124710065693 from 9 Aug 2021; only a part of the original image is showed here)

On the other hand, this framework implies that it is possible to infer the demographic characteristics of pathogen populations, such as their rate of growth, using statistical models based on coalescent theory that allow to directly link patterns of genetic diversity to ecological processes, such as changing population size and population structure (Figure 1.21) (132).
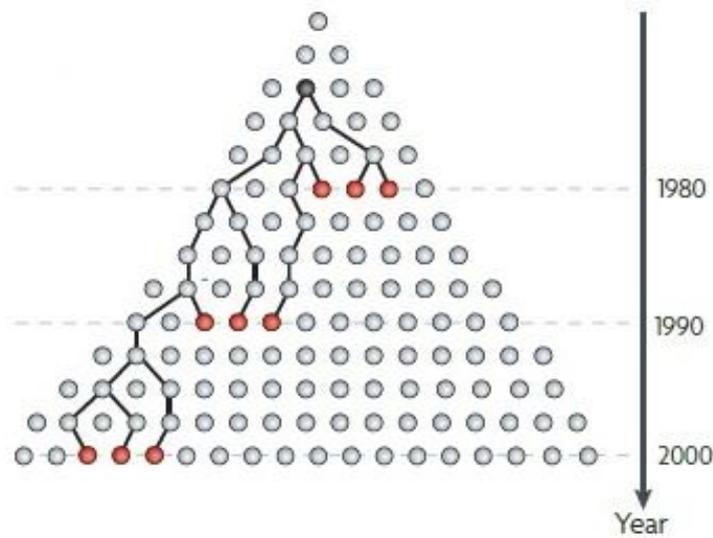
Figure 1.21 | Coalescence method. This approach is typically used to infer past rates of epidemic growth from sampled viral sequences. Each circle represents an infection, and circles on the same row occur during the same period. The increasing width of each row therefore reflects the growth of the epidemic through time. Starting from the sampled infections (red), the sampled lineages (black lines) can be traced back through unsampled infections (grey) to the common ancestor (black circle). The rate at which the sampled lineages merge or coalesce depends on population processes such as population dynamics, population structure, selection, and recombination (only change in population size is represented here). (Original image from Oliver Pybus and Andrew Rambaut, Nature Reviews Genetics 2009; Licensed under © 2009 Macmillan Publishers Limited. All rights reserved, but reprinted by permission from Springer Nature under the license number 5124710065693 from 9 Aug 2021; only a part of the original image is showed here)

Such changes in branching structure and effective population sizes (Ne) across time can then be plotted using a variety of graphical methods. While parametric models imply assumptions about past patterns of population growth – e.g., exponential, or logistic growth models – the skyline plots are particularly useful as they provide a nonparametric estimate of the change in viral effective population size, thereby outlining the demographic history of that population through time. More sophisticated coalescent approaches allow combining different models of population growth, such as constant, exponential and logistic growth (133).

As an example, an analysis of HCV in Egypt using coalescent-based methods reconstructed a mid-twentieth century explosion in transmission that was linked to widespread unsafe injection during campaigns against schistosomiasis (134).

1.2. Hepatitis B origin

Molecular-clock-based analysis has been used to reconstruct the evolution of HBV as well as dating its origin, but the results have proven to be still unclear and controversial (135-137). A theory from the 90s suggests an Old World origin of HBV, and that it was spread following prehistoric human migrations over 100,000 years ago (138). In 2013, Paraskevis and colleagues provided some support to this theory showing that HBV "jumped" into humans between 22,000 and 47,100 years ago from an unknown source and proposed that humans then went on to infect non-human primates (NHP) in Africa, Asia, and the New World (139).

By the end of the 1990s, a New World origin was proposed for HBV due to the discovery of a phylogenetic co-segregation between the New World HBV genotypes F and H and Wolly Monkey HBV in a close relationship to the Old World NHP and human HBV variants (140). It was suggested the dissemination of HBV into the Old World may have occurred around 500 years ago. However, this theory is difficult to reconcile with recent evidence that HBV was already circulating in 16th century in East Asia due to the discovery of the virus in a Korean mummy (141) with its origin dated to at least 3000 years ago in that region. A third theory suggests a co-speciation of HBV in NHP hosts because of the proximity between the phylogeny of Old and New World NHP and their HBV variants (142). More recently, HBV was found in Medieval and Neolithic human remains from Europe that seem to date back nearly 7000 years ago (143).

None of these theories is proven, and the reality probably lies in a combination of more than one of these theories with co-species evolution within birds, rodents, and bats, followed by a series of cross-species transmission events to explain the close relationship between human and nonhuman primate HBVs observed today (137). However, as more HBV sequences are found in ancient remains (139, 141, 143, 144), the theory that this virus emerged many thousands of years ago becomes more likely (145).

1.2.1 Genotypes and subgenotypes possible origin and spreading

The origin of the circulating HBV genotypes and subgenotypes and their dispersion have been object of study by several research groups. Most of the studies about HBV/A subgenotypes have been relating their emergence to Africa, but HBV/A2, which is highly prevalent in Europe and North America, has been the exception (67). Its time-scaled

phylogeny suggested a relatively recent origin of the currently circulating strains, probably during the first decades of the 20th century, also showing it seems to be the only subgenotype of HBV/A that has differentiated itself in the western world, possibly in the United States of America (USA) or Europe (146). Its wide dispersion is thought to be related with high-risk sexual behavior, specially between MSM (146). As for HBV/A1, Kramvis and Paraskevis traced back its origin to Southern Africa (147), while another study suggested East Africa as the origin of the virus (146). In any case the most probable dispersion pathway outside this continent seems to have been the African slave trade period to American and Asian continent (147)

More recently, with information obtained from ancestral HBV sequences, the hypothesis that the A genotype may have originated in a region close to Africa and Europe, such as the Middle East/Central Asia gained visibility, from where it may have dispersed to both continents Europe and Africa (90).

Phylogeography of HBV/D is also widely studied. Some hypothesis indicated North Africa/Middle East (90), Southern Europe, Central/Eastern Europe, Syria or Martinique (91) as its putative origins. In Americas there are several studies on HBV/D subtypes origins (91, 148-151), and it was possible to hypothesize that dispersal pathways of HBV/D3 in the Americas were complex including multiple introductions from different geographic regions. Brazil is the largest country in the Southern Hemisphere corresponding to almost half of the area of South America and it seems that most of HBV/D3 from Southern Europe during the mass European emigration from different countries encouraged by the Brazilian government campaigns (152). About 13 million Europeans went to Latin America, mostly Italians, Spaniards and Portuguese (153), but also from Germans, and people from some Asiatic countries such as Japanese, Lebanese and Syrians (152). Phylogeographic analysis also suggested Brazil as the putative origin of most dispersal pathways occurring inside the American continent (150, 152). As for HBV/D4, the spatiotemporal reconstruction suggested that Martinique is the most likely origin of this subgenotype in the American continent, but this result was considered incompatible with epidemiological and historical data of HBV in the Americas (91). A clustering effect of HBV/D4 Brazilian sequences was previously described and a single introduction of HBV/D4 in Brazil had been proposed by the authors with a probable time of introduction of 1848 (1062 to1946) but undefined origin (91).

HBV/E has been assumed to have a more recent origin probably in the present region of Nigeria around the last 130 to 200 years, which is supported by its low genetic variability characteristic of a shorter evolutionary history (94, 154). Despite the low genetic variability, HBV/E is hyperendemic throughout West Africa, and it has spread rapidly in countries along the West African coast from Guinea to the Central African Republic and subsequently towards Eastern and Southern countries, such as Sudan, Angola, and the Democratic Republic of the Congo (94, 154). As reviewed by Andernach *et al* in 2009, an explanation for this phenomenon would be iatrogenic transmission in mass-injection campaigns with unsafe injections by the colonial governments (155). The period between the 1880s and 1930s was described as the most significant expansion period of this genotype (94). The absence of HBV/E in descendants of African slaves in South America was assumed to indicate that this genotype could not have been present in the West African population in the period of slave trade (94, 148, 149). The absence of the HBV/E in the South American continent also indicates that Europeans who emigrated to the Americas did not introduce it there.

The MRCA of HBV/C was traced back to the early 1900s in China, where it eventually diverged into two major lineages during the 1930s and 1960s giving rise to distinct epidemic waves spreading exponentially to other East Asian countries and the USA (156). The evolutionary analysis using complete genome sequences estimated an older median ancestral age for the Brazilian HBV/F2a compared to the Brazilian HBV/F1b and HBV/F4 subgenotypes, suggesting that HBV/F2a represents the original native HBV of Brazil. The phylogeographic patterns suggested a north-to-south flow of HBV/F2a from Venezuela to Brazil, whereas HBV/F1b and HBV/F4 strains appeared to have spread from Argentina to Brazil (157).

The HBV/G has a possible origin in North America dating back to 1855 (95% highest posterior density interval [95% HPD]: 1778 – 1931) and it may have been further disseminated to other continents (South and Central America, Europe, Asia, and Africa) more than one century later (around the 1970s). The viral population demonstrated constant spread from 1855 to the 1980s, growing faster from the 1980s to the 1990s, which may be related to highly sexually active groups and injecting drug users, reaching a plateau after the 2000s (96). Overall HBV/H time to the most recent common ancestor (tMRCA) dated back to 1933 (95% HPD: 1875 – 1957) with a very probable origin in

Mexico and posterior dissemination to other American and Asian countries (66).

In recent years several ancestral HBV sequences were discovered in ancient human remains and a phylogenetic study involving all of them and the extant HBV human genotypes revealed genetic affinities between a sequence from Bronze age isolated in Russia and genotype A (145). The basal position of that ancestral sequence within genotype A clade led to a proposed origin of this genotype in central Asia (90, 144). The same basal pattern between a sequence, isolated from a male´s remains from Saka people of around 297 B.C in Central Asia, and the extant genotype D also suggest it represents an ancestral sequence or extinct subgenotype of the present D genotype (145). This collective analysis including all available ancient HBV (aHBV) sequences also helped identify several recombination events never previously recognized, that contributed to the evolution of HBV genotypes. For example, a signal of recombination between a Neolithic and an Iron age HBV sequence is present in all extant subgenotypes of HBV/A suggesting these subgenotypes emergence is an ancient event. A common ancestor of the extant HBV/A and HBV/I also revealed this same ancient signal of recombination, which indicates that despite I genotype has been identified recently may also have an ancient origin as proposed to genotype A. The extant genotypes E and G also revealed the presence of an Iron age HBV sequence, suggesting a more ancient origin than was described until now. Interestingly, some aHBV sequences revealed that part of their genome has a great genetic affinity with the extant HBV/E reinforcing the hypothesis of an ancient pre-existence of genotype E. In addition, the presence of gibbon HBV (presently restricted to Asia) sequence was also found in a Neolithic aHBV from Europe which might be a resultant of ancient interspecies HBV transmissions.

The search for aHBV genomes in archaic human remains is allowing to better understand the distribution and dispersal of HBV through times.

1.3. HCV origin

The origin of HCV as a human pathogen is still obscure. The closest relatives of HCV are equine and canine hepaciviruses (158, 159). Some authors hypothesized that HCV evolved from a horse-to human transmission event (160, 161), others suggested that HCV originated in relatively recent times from one or multiple cross-species transmission events from a still to be defined species (158, 159, 161). However, its strict species

specificity, as well as the ability of HCV to persist lifelong in humans, led to the alternative hypothesis whereby HCV related viruses have been infecting humans and other primates throughout their evolutionary history (161, 162). Some studies provided estimates of the tMRCA of HCV genotypes in a range between 200 and 1000 years ago, with one study indicating that HCV origin may date 2000 years back (162-167). The tMRCA of equine and canine hepaciviruses was estimated to be recent, dating around 1800 CE (159), ruling out the possibility that the human virus derived from a cross-species transmission event from horses (168).

On the other hand, a recent study estimated the tMRCA of the seven HCV genotypes to be around 3359 years ago with a confidence interval (CI) of 5221 to 3192 (168). This time frame of HCV origin and the geographic distribution of the endemic strains are not consistent with the possibility that HCV dispersed with humans following the major out-of-Africa colonization routes across the Old World (65000 – 45000 years ago) (169). Instead, it may suggest that the (possible) zoonotic transmission and subsequent spread in human populations occurred in a time frame when long-distance trade routes were being established.

Recently, a phylogenetic reconstruction study showed that root line HCV/Non-Primate Human Virus split from other listed lines 985 - 2613 years ago, probably in East Africa (170).

HCV may have established in humans a long time ago, but its worldwide spread seems to have started recently, in the 1930s – 1940s (162).

1.3.1 Genotypes and subtypes origin and spreading

A small number of "epidemic" subtypes (1a, 1b, 3a, and 2a) account for the overwhelming majority of HCV infections worldwide (75, 162).

A detailed phylogenetic analysis of the HCV determined that the common ancestor of all of the seven HCV genotypes dated back to near 780.86 with a 95% HPD from 592.15 to 1021.34 years ago (166), while subtypes were estimated to have emerged in the last 300 years (171). However, more recently the tMRCA of the seven HCV genotypes was estimated to be 3359 years ago with a CI from 5221 to 3192 (168), with the deepest tMRCA obtained for HCV genotype 6 (HCV/6) (endemic in South-East Asia), with an origin dating at least 2000 years ago. HCV/6 tMRCA had previously been dated to 1100

- 1350 years ago but the place of origin was not clear (172). Possibly, this indicates an Asian origin of extant strains (168). The tMRCAs of HCV genotype 3 (HCV/3) (endemic in South Asia) and of the ancestor of HCV genotypes 1 (HCV/1) and HCV genotype 4 (HCV/4) (both widespread in Central Africa) were estimated to be 800 before common age (BCE) – 700 common age (CE) range (168). Some authors had previously speculated that the ancestral HCV/3 strains may have been brought to South Asia from Africa by land and/or across the sea to result in its indigenous circulation in that region. The spread was estimated to have occurred in the era after Vasco da Gama had completed his expeditions by sailing along the eastern coast of Africa to India. However, before this era, Arabian slave trading from Africa to the Middle East and South Asia, may have mediated the earliest spread of HCV/3 (166). The divergence time of HCV/3 was near 457.81 with a 95% HPD from 350.62 to 587.53 years ago (166).

HCV/4 seems to have been originated in central Africa. However, some authors point to 1539 (173) while other estimate that the HCV/4 common ancestor existed around 1733 (1650–1805) (174). It seems that multiple lineages from central Africa have been exported to north Africa since 1850, including subtype 4a which dominates the epidemic in Egypt (174).

It was suggested that HCV/1 arose, around 675 years old back in Nigeria (175) and that subtypes 1a and 1b diverged from common ancestors 135 and 112 years ago, respectively (176). Phylogeographic analysis of all available NS5B sequences suggests that HCV subtypes 1a and 1b disseminated from the developed world to the developing countries and the transmission of these subtypes ''exploded'' between 1940 and 1980, with the spread of 1b preceding that of 1a by at least 16 y. It probably coincides with the vast increase in transfusions and unsafe therapeutic injections, whereas the expansion of HCV 1a is more strongly associated with the increase of intravenous drug users after 1960 (177).

Forni and colleagues indicated an origin, for HCV genotype 2 (HCV/2) (endemic in West Africa), around 430 years ago (168), while previous molecular clock analysis dated the common ancestor of HCV/2 in Guinea-Bissau to 1470 (1414 – 1582) with the spatial analysis showing the exportation from the west African coast to Cameroon that took place over several centuries and suggesting that infections moved to the New World via Benin–Ghana (178). It is interesting to note that, specifically, the epidemic subtypes 2a and 2c,

appear to have both originated from the Benin – Ghana area, with slave trade from 17th and 18th centuries playing a historical role in the global dissemination of these subtypes and HCV/2 in general (178).

AIM

The aim of this research was to perform the first characterization of the evolutionary history of HBV and HCV in Portugal.

OBJECTIVES

The main objectives of this research study were to investigate:

1. the epidemiological history of HBV and HCV in Portugal.
2. the origin and routes of spread of the most frequent subgenotypes and subtypes of HBV and HCV, respectively.
3. the presence of natural NS5A RAS in HCV/1a clades strains.

We proposed to address the following main research questions:

For objective 1

a) Which are the most prevalent genotypes/subtypes circulating in our population?

b) Which factors may have contributed to their entrance in Portugal?

For objective 2

a) Which countries were the major contributors to HBV and HCV introduction in Portugal?

b) Which was the role of Portugal in the spread of each virus?

For objective 3

a) Is there any association between HCV/1a RAS presence and clade type?

b) Is there any association between HCV/1a RAS presence and treatment outcome?

c) Is there any association between HCV/1a RAS presence and any characteristics of the population studied?

MATERIALS AND METHODS

2.1 Phylogeography of HBV

2.1.1 Study participants and ethics statement

In this retrospective study we used plasma samples collected for diagnostic purposes between 2005 and 2012 from chronic HBV Portuguese (PT) patients assisted in Department of Gastroenterology and Hepatology of Santa Maria Hospital, Medical School of Lisbon, Portugal. One hundred and thirty samples were selected for sequencing based on sample volume, the detection of viral load and the availability of information on age and gender. Informed consent was obtained from all subjects. This study was approved by the Institutional Review Board of Santa Maria Hospital (ref. 600/15 of January 7th, 2016) and followed all ethical principles of the Declaration of Helsinki, as revised in 2013.

2.1.2 HBV DNA extraction, *pol* gene amplification and sequencing

Viral nucleic acid was extracted from 200 µl of plasma using the QIAamp DNA Blood Mini Kit (Qiagen, Germany) according to manufacturing directions. A 943 bp fragment of the *pol* gene region was amplified by nested polymerase chain reaction (PCR) using the forward (FW) and reverse (RV) primer pairs described in table 2.1 (See details of the protocols in annex 1).

Table 2.1 HBV PCR primers sequences, position in the reference sequence NC_003977 and fragments size.

Primer	Sequence	Position (bp) (In NC_003977)	PCR round	Fragment Size
FW_Outer	5'-ATCCTCACAATACCGCAGA-3'	229-247	First	1056bp
RV_Outer	5'-AGGAGTTCCGCAGTATGG-3'	1285-1268		
FW_Inner	5'-AGACTCGTGGTGGACTTCTCT-3'	252-272	Second	943bp
RV_Inner	5'-GCGTCAGCAAACACTT-3'	1195-1180		

The amplification process always included positive and negative controls. Following PCR Product Purification with JetQuick PCR Product Spin Kit (Genomed, Germany), amplified DNA products were sequenced on the automated sequencer ABI PRISM 3100-Avant Analyser (Applied Biosystems, USA) using the second PCR round primer pair and BigDye terminator v. 3.1 cycle sequencing kit (Applied Biosystems, USA) (See details of the protocols in annex 1). The sequences were analyzed using Chromas Pro.1.7.6 software (Technelysium, Pty, Australia).

2.1.3 HBV genotyping

Multiple sequence alignment was performed using the ClustalW algorithm, as implemented in the software ClustalX 2.1 (179) and edited using Sea View (180). Genomic sequences were genotyped using the Geno2Pheno [HBV] online tool (https://hbv.geno2pheno.org/). The reliability of the Geno2Pheno results was confirmed latter by analyzing the consistency of its results with the results of the phylogenetic analysis.

For the phylogenetic tree reconstruction, HBV reference sequences belonging to the different genotypes and subgenotypes were retrieved from NCBI Blast Nucleotide (https://blast.ncbi.nlm.nih.gov/). The PT HBV sequences, and the reference sequences were aligned using MAFFT algorithm (181) and edited using Sea View (180). The IQtree 1.6.11 software (182) was used to reconstruct a ML tree under the best fit evolutionary model [GTR+F+R4, selected using Model Finder module (183) and the Akaike information criterion (184)], with 1,000 bootstrap replicates. Final visualization and annotation of the ML tree was performed with FigTree v1.4.4 (http://tree.bio.ed.ac.uk/software/figtree/).

2.1.4 Genetic Distances

Evolutionary distances were estimated using the Kimura 2-parameter evolutionary model incorporating rate variation among sites that was modelled with a gamma distribution (shape parameter = 1). Sequences were grouped according to genotypes identified, as explained in the previous section. Between group and within group evolutionary distances were calculated as the average pairwise evolutionary distance between all pairs of sequences belonging to distinct groups (between groups distance) and the average pairwise evolutionary distance of all pairs of sequences within each group (within group distance), respectively. Standard error estimates were obtained using a bootstrap procedure (1000 replicates). All analyses were conducted in MEGA7 (185).

2.1.5 Reference sequences selection for spatiotemporal analysis

Reference sequences for the spatiotemporal analysis were selected using the National Center for Biotechnology Information (NCBI) Nucleotide Basic Local Alignment Search Tool (BLAST) (https://blast.ncbi.nlm.nih.gov/) algorithm. There were 9489 HBV complete

genomes and 6561 HBV *pol* gene sequences available. Each of the 130 PT HBV sequences was submitted to NCBI nucleotide blast, selecting a maximum target result of 50 sequences in the algorithm parameters. Thus, for each of the Portuguese sequences, we retrieved the 50 most similar HBV sequences available in Genbank, making a total of 6500 HBV sequences for the reference group. Multiple sequence alignment was performed on this group of 6500 sequences using ClustalW, as implemented in the ClustalX 2.1 software (179). Duplicate sequences were removed from this dataset with a script in Python v3.7.0 (https://www.python.org/downloads/release/python-370/), reducing the data file to 687 HBV reference sequences. From this group, 286 sequences without collection date and/or country of origin were excluded, reducing the data set to 401 reference sequences. An effort was made to keep the temporal and spatial information as wide as possible in the reference sequence dataset, diversifying as much as possible both the year of harvest and the country of origin of the reference sequences. A manual search in the NCBI was carried out, selecting the sequences by genotype and year of harvest to increase the representativeness of the 60s, 70s and 80s for most genotypes. We also had to increase the representativeness of the origin countries of A1 and D4 subgenotypes, to ensure that Brazil was not falsely over-represented in alignment since the result of the initial BLAST nucleotide resulted in near 35% of sequences from that country. With this approach, the final dataset consisted of 453 HBV reference sequences spanning from 1963 to 2018 (Annex 2_Table A2), and 130 PT HBV sequences.

2.1.6 Spatiotemporal evolutionary dynamics

To minimize the effect of convergent evolution, nine codon positions reportedly associated with treatment selective pressure were removed from the alignment. Time-scaled phylogeny, evolutionary rate and phylogeography were estimated using a Bayesian MCMC method implemented in the BEAST v1.10.4 (186), with the GTR+F+R4 nucleotide substitution model that was selected using Model Finder module (183) and the Akaike information criterion (184) within iqtree 1.6.11 software (182).

Eight different evolutionary models were tested: strict vs. relaxed molecular clock, each one combined with either parametric (expansion Growth and exponential Growth) and non-parametric (Bayesian Skygrid and Bayesian Skyline) demographic priors and with a continuous Cauchy RRW phylogeography model. In each case, two independent MCMC

chains were run for 300 million generations (sampled every 30.000 steps) and both Log files were combined using Logcombiner to ensure convergence and Effective Sample Size (ESS) values > 100. Our data only reached convergence and showed ESS values > 100 when using an uncorrelated lognormal relaxed molecular clock model under a Bayesian skyline plot (BSP), a non-parametric piecewise-constant model, as coalescent priors.

Uncertainty of parameter estimates was assessed by calculating the 95% HPD values after excluding the initial 10% of the run using TRACER v1.7.1 program (187). Maximum clade credibility (MCC) trees were summarized from the posterior distribution of trees with TreeAnnotator included in BEAST package and visualized and annotated with FigTree v1.4.4 (http://tree.bio.ed.ac.uk/software/figtree/) to show node support values, the posterior state probability (PSP) and median time of MRCAs with 95% HPD for each common ancestral node. Bifurcating nodes with PSP greater than 0.90 were considered statistically well supported.

The programs BEAUti, BEAST, Tracer, and TreeAnnotator were used according to the tutorials of beast community website (188) and following guidelines previously described for continuous phylogeographic inference analysis (189) .

2.2 Epidemic history of HCV genotypes and subtypes in Portugal

2.2.1 Study participants and ethics statement

This was a retrospective observational cross-sectional study of consecutive HCV-infected patients followed-up between November 2007 and July 2009 at the Department of Gastroenterology and Hepatology of Santa Maria Hospital, Medical School of Lisbon, Portugal. Eligibility criteria include having a diagnosis of HCV infection based on detectable viral load and available information on age and gender. Stored plasma samples were retrieved from 230 patients. Informed consent was obtained from all subjects. The study was conducted in accordance with the Declaration of Helsinki, as revised in 2013, and was approved by the Institutional Review Board of Santa Maria Hospital (ref. 245/15 of July 30, 2015).

2.2.2 Extraction of viral RNA, amplification, and sequencing of HCV NS5B

Viral RNA was extracted from plasma with the QIAmp Viral RNA Mini kit (Qiagen, Germany) and a 372 bp region of NS5B gene corresponding to the Okamoto region was amplified using Titan One Tube RT-PCR System (Roche Diagnostics, Switzerland) with the following FW_OKA1 and RV_OKA2 primers described in table 2.2 (See details of the protocols in annex 3_Part A).

Table 2.2 HCV PCR primers sequences, position in the reference isolate H77 and fragment size.

Primer	Sequence	Position (bp) (In H77: AF009606)	Fragment Size
FW_OKA1	5′-CCCGCTGYTTTGACTCVACNGT-3′	8264–8285	372bp
RV_OKA2	5′-CCTRGTCATAGCCTCCGTGAA-3′	8636–8616	

Negative and positive controls were included in all amplification procedures. Amplicons were purified with JetQuick PCR Product Purification Spin Kit (Genomed, Germany) and sequenced on the automated sequencer ABI PRISM 3100-Avant Analyser (Applied Biosystems, USA) using the same primers used for amplification and BigDye terminator v. 3.1 cycle sequencing kit (Applied Biosystems, USA) (See details of the protocols in annex 3_Part A). The sequences were analyzed using Chromas Pro.1.7.6 software (Technelysium, Pty, Australia).

2.2.3 HCV genotyping and subtype assignment

The 230 nucleotide sequences were aligned with MAFFT algorithm (181) as implemented in the HIV Align tool hosted at Los Alamos HIV Database (https://www.hiv.lanl.gov/) and edited using SeaView (180) HCV subtyping and clade genotype 1/subtype a (HCV/1a) lineages (clade I and clade II) confirmation were both performed using geno2pheno[HCV] (190) (http://hcv.geno2pheno.org/). The consistency of these results was analyzed by comparing the results with the phylogenetic ML tree that was reconstructed under a GTR + Γ nucleotide substitution model and 1000 bootstrap replicates as implemented in raxMLGUI v1.5 (191).The HCV genotyped sequences were deposited in GenBank under accession numbers MG821636-MG821865. The accession numbers for the additional GT4d sequences (see section 2.2.4 below) include: FN401072, FN401080, FN401090, FN401092, FN401095, FN401120, FN401132, FN401142, FN401146, FN401154, FN401157, FN401177, FN401182, FN401193.

2.2.4 Bayesian evolutionary analysis

Time-scaled Bayesian phylogenies were reconstructed using the Bayesian Evolutionary Analysis by Sampling Trees software package (BEAST v1.10.4) (186) for HCV subtype datasets containing at least 10 sequences with information about date of sampling, spanning the sampling years 2007 – 2009 (subtypes 1a, 1b, 3a, 4a and 4d). To improve the evolutionary estimates of genotypes 4/subtype d (HCV/4d), 14 HCV/4d sequences (sampling years 2008 – 2009) previously reported in Portugal (103) were retrieved from the Los Alamos database. The SRD06 two-codon partitioning model (192), the HKY + 4Γ nucleotide substitution model, and an uncorrelated lognormal relaxed molecular clock (UCLD) with a BSP coalescent tree prior were selected for the Bayesian analysis. To estimate the date of the MRCA and the population growth dynamics of the HCV epidemic in Portugal, we specified a Normal prior distribution on the UCLD mean clock rate parameter (0.001 ± 0.0001 s/s/y) (174, 193, 194). For each dataset, three independent MCMC chains were run for 100 million generations with states sampled every 10,000 generations. Log files were combined using Logcombiner to ensure sufficient convergence (ESS ≥ 100) as monitored in Tracer v1.7.1 (187) with 10% of posterior samples discarded as burn-in. MCC trees were summarized using tree annotator and tree visualization was performed in Figtree v1.4.4 (http://tree.bio.ed.ac.uk/software/figtree/).

2.2.5 Statistical Analysis

Categorical variables were analyzed using the chi-squared test or Fisher's exact test. The analysis was 2-tailed and statistical significance was defined as p-values < 0.05. Statistical analyses were performed using the SPSS software Version 22.0 (IBM Corp, Chicago, Armonk, NY, USA).

2.3 Evolutionary history and phylogeography of Clades I and II of HCV subtype 1a in Portugal

The reason why it was decided to focus this part of the study only on HCV/1a, and particularly on clades I and II, is related to the results obtained for the "Epidemic history of HCV genotypes and subtypes in Portugal" and the explanation is given in the beginning of section "Evolutionary history and phylogeography of Clades I and II of HCV subtype 1a in Portugal" of chapter 3 - Results.

2.3.1 Study participants selection and ethics statement

To perform this part of the study, HCV/1a samples were selected from a source population used to study baseline NS5A and NS5B Resistance Associated Substitutions (RAS) in naïve HCV/1a and HCV/1b persons in the context of another study. Between 2015 and 2017, blood samples from HCV/1a and HCV/1b infected patients, naïve for DAA therapeutics and belonging to the list of patients waiting for DAA regimen to be initiated, were collected at Centro Hospitalar de Lisboa Ocidental (CHLO). After collecting samples to determine baseline resistance in the virus genome, all patients began treatment with Sofosbuvir and Ledipasvir combination. Plasma was obtained by blood sample centrifugation and stored at -80ºC. From these initial group of 121 collected samples, some were excluded from the resistance analysis study due to the lack of clinical data information about previous therapeutics, due to following loss of patient after the treatment with DAA and consequently unknown treatment outcome and finally due to lack of sequence quality to analyze baseline resistance. The remaining patients, a group of 83 individuals, had samples successfully sequenced for both NS5A and NS5B regions. Clinical data such as age, gender, HCV mono or coinfection with HIV and/or HBV, viral load, IL28B and result after DAA treatment of 12 or 24 weeks were available for each sample. Subtype had been previously determined using the VERSANT® HCV Genotype 2.0 Assay Line Probe Assay (LiPA) from INNOGENETICS/Siemens Healthineers (Ghent, Belgium). Among the 83 samples from the resistance baseline study, 58 were classified as HCV/1a.

To the present part of this study, only these HCV/1a samples were selected once the purpose is to analyze the evolutionary history of clades I and II of that HCV subtype. This study was approved on 6 September 2017 by the Ethical Committee for Health of

CHLO with the RNEC Registry Number: 20170700050 and followed all ethical principles of the Declaration of Helsinki.

2.3.2 HCV RNA extraction and amplification of NS5 region

HCV nucleic acid was extracted from 500 µL of plasma previously conserved at -80ºC, using the bioMérieux's automated NucliSENS EasyMAG system v2.0, with silica, (Boxtel, The Netherlands), according to the manufacturer's protocol. To cover the complete NS5A and NS5B region, a 3687 bp fragment of the viral genomic region was amplified by nested PCR using the FW and RV primer pairs described in table 2.3 (See details of the protocols in annex 3_Part B).

Table 2.3 HCV PCR primers sequences, position in the reference sequence isolate H77 and fragments size.

Primer	Sequence	Position (In isolate H77)	PCR round	Fragment Size
FW1_PCR	GTCGTCACBAGYACCTGG	5311 - 5328	First	4005bp
RV1_PCR	GAGACASGCTGT GATAWATG	9297 - 9316		
FW2_PCR	GTGGTCATHGT GGYAGG	5386 - 5403	Second	3687bp
RV2_PCR	CCCTATTGATYT CACCTGG	9055 - 9073		

The amplification process always included positive and negative controls. Following PCR Product Purification with ExoSAP-IT™ PCR Product Cleanup (Applied Biosystems, USA), amplified DNA products were sequenced on the automated sequencer ABI PRISM 3130xl Genetic Analyzer (ThermoFisher Scientific, USA) using the BigDye terminator v. 3.1 cycle sequencing kit (Applied Biosystems, USA) and the primers described in table A5.1 at the end of annex 3 (See details of the protocols in annex 3_Part B).

For each sample, nucleotide sequences obtained with each sequencing primer were joined into a single consensus sequence resulting from the overlapping of all DNA fragments using ChromasPro software v1.7.6. (Technelysium Pty Ltd., Australia). All sequences obtained in this study were submitted to the REGA HCV database from Katholieke Universiteit Leuven (KU LEUVEN) (195).

2.3.3 HCV genotype and subtype assignment

The 83 nucleotide sequences from the source population of samples were aligned with MAFFT algorithm (181) as implemented in the HIV Align tool hosted at Los Alamos HIV Database (https://www.hiv.lanl.gov/) and edited using SeaView (180). Using the complete NS5A sequence of each sample, HCV genotypes/subtypes were obtained through phylogenetic analysis. For the phylogenetic tree reconstruction, the first step was to retrieve HCV subtype reference sequences belonging to the different genotypes and subtypes from NCBI Blast Nucleotide (https://blast.ncbi.nlm.nih.gov/). The PT HCV sequences and the reference sequences were aligned using MAFFT algorithm (181). The IQtree 1.6.11 software (182) was used to reconstruct a ML tree under the best fit evolutionary model [TIM2+F+I+G4, as selected by Model Finder (183) and the Akaike information criterion (184)], with 1,000 bootstrap replicates. Final visualization and annotation of the ML tree was performed with FigTree v1.4.4 (http://tree.bio.ed.ac.uk/software/figtree/).

The results were compared with those obtained through LiPA assay and Geno2Pheno [HCV] 0.92 (190) .

Later, this same tree reconstruction was repeated only with the 58 HCV/1a sequences to distinguish in more detail the clade I and clade II.

2.3.4 Characteristics of the selected HCV/1a population and statistical analysis

Demographical and clinical data of the 58 HCV/1a samples were stratified by clade type. The association between each variable and belonging to clade I or clade II was statistically analyzed. The Chi-square test was used to analyze categorical variables under a Fisher exact test as some cross-table cells had less than 5 expected values. Previously to t-student analysis to compare means, the distribution of continuous variables was analyzed to test its approximation to a Poisson curve and the equality of variances was compared through a Levene's test. Statistical significance α was 5% and statistical tests were two tailed. The SPSS software v.26 (SPSS Inc., Chicago, IL) was used to perform analyses.

2.3.5 Reference sequences selection for spatiotemporal analysis

Given that phylogenetic analysis classified the PT HCV/1a sequences in two different clades (I and II), two datasets were produced containing the PT and HCV reference sequences for each clade. Reference sequences for the spatiotemporal analysis were selected in the NCBI Nucleotide Blast (https://blast.ncbi.nlm.nih.gov/) algorithm, using an approach like the one described in part 1 of this chapter, for HBV sequences. Sequences without collection date and/or information about country of origin were excluded from the selection groups. As a result, the final datasets consisted of 165 HCV reference sequences plus 15 PT HCV sequences for clade I and 194 HCV reference sequences plus 43 PT HCV sequences for clade II. Each clade dataset was aligned in MAFFT (181) and after editing in the SeaView (180) both datasets had 1314 bp long.

2.3.6 Spatiotemporal evolutionary dynamics

Using TempEst v1.5 (196), a regression of root-to-tip genetic distances against year of sampling, using a heuristic residual mean squared function, was carried out to investigate the temporal signal of clade I and clade II sequence datasets.

To investigate the origin and dispersal route of each HCV/1a clade a Bayesian MCMC analysis, implemented in BEAST v1.10.4 (186), was carried out with a HKY nucleotide substitution model that was selected using Model Finder module (183) and the Akaike information criterion (184) within IQtree 1.6.11 software (182), a SRD06 two-codon partitioning model (192), and an UCLD with a BSP (non-parametric piecewise-constant model) coalescent specified prior. As described previously (193), we also specified a Normal prior distribution on the clock rate (0.001 ± 0.0001 substitutions/site/year). This approach was performed over continuous sampling locations analyzed with a Cauchy RRW model. The MCMC chains were run in triplicate for 2×10^8 and 2.5×10^8 generations for clades I and II, respectively, to ensure convergence and ESS values > 100. The convergence of MCMC chains was monitored in Tracer v1.7.1 (187). Maximum clade credibility trees were summarized using TreeAnnotator from BEAST package and tree visualization was performed in Figtree v1.4.4 (https://tree.bio.ed.ac.uk/software/figtree/), as well as annotations to show node posterior probability values, the PSP and median time of MRCAs with 95% HPD for each common ancestral node. Bifurcating nodes with posterior probability greater than 0.90 were considered statistically well supported.

2.4 Characterization of NS5A RAS from the two clades of HCV/1a

2.4.1 Study participants and ethics statement

This part of the study follows on from section 2.3 and was carried out with the same samples. For this reason, the approval process was the same of the previous section (6 September 2017 by the Ethical Committee for Health of Centro Hospitalar de Lisboa Ocidental (CHLO) with the RNEC Registry Number: 20170700050) and followed all ethical principles of the Declaration of Helsinki.

2.4.2 Resistance Profile Analysis

The nucleotide sequences of NS5 region of the 58 sample were analyzed online in Geno2Pheno [HCV] 0.92 (190) to obtain the baseline resistance analysis profile for each patient. All RAS and the respective resistance profile to each antiviral was registered in a database created to the purpose of the study.

2.4.3 Data analysis

The presence of RAS and the treatment outcome (sustained virological response and relapse) of the 58 HCV/1a samples were both stratified by clade type in a first approach, and then were also stratified by subclades of clade I. The association between each variable and belonging to clade I or clade II (or subclades) was statistically analyzed as described in the next section.

Each RAS was annotated in the respective strain (branch) of the clades and subclades phylogenetic tree, to facilitate the visualization of their frequency in each of these groups.

2.4.4 Statistical Analysis

The Chi-square test was used to analyze categorical variables under a Fisher exact test as some cross-table cells had less than 5 expected values. Statistical significance α was 5% and statistical tests were two tailed. The SPSS software v.26 (SPSS Inc., Chicago, IL) was used to perform analyses.

RESULTS

3.1 Phylogeography of HBV in Portugal

3.1.1 HBV genotype A was the most prevalent in Portugal

Most (n=97, 74.6%) of the patients included in the study were male with a median age of 43.6 years. Median age of females (n= 33, 25.4%) was 41.2 years.

According to our genotyping phylogenetic analysis (Figure 3.1), which agrees with Geno2Pheno genotype classification, HBV/A was the most prevalent genotype (n=54, 41.5%), followed by HBV/D; (n=44, 33.8%), and HBV/E (n=32, 24.6%)].
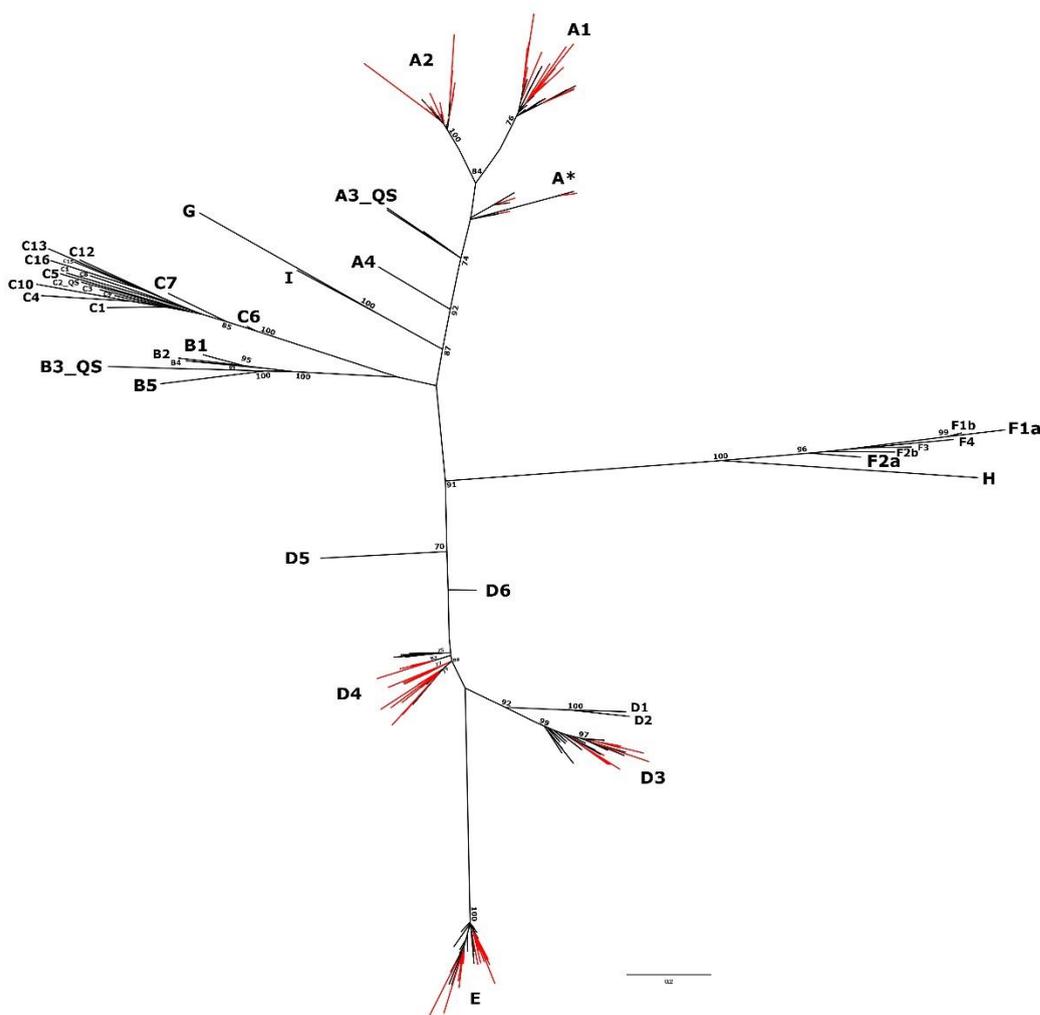
Figure 3.1 | ML phylogenetic tree reconstruction showing the classification of 130 Portuguese HBV samples into genotypes and/or subgenotypes. The tree was reconstructed using IQtree 1.6.1 software under a GTR+F+R5 evolutionary model with 1000 bootstraps replicates. PT HBV strains are represented by the red external branches and reference strains by the black external branches. Confirmed genotypes A, B, C, D, E, F, G, H and I are represented along the tree, each one showing most of their subgenotypes. A* represents a group of strains classified as belonging to genotype A but without any subgenotypes assigned. Bootstraps above 70% are annotated in the branches. The scale in the bottom right represents the number of substitutions per site.

Subgenotypes HBV/A1 and HBV/D4 were almost equally prevalent with 23.1% (n=30) and 22.3% (n=29), respectively, followed by HBV/A2 with 16.2% (n=21) and HBV/D3 with 11.5% (n=15). Three sequences (n=2.3%) were classified as genotype A without further subgenotype classification.

Genetic distances between HBV genotypes and subgenotypes present in Portugal were >7.5% for genotypes and between 4% - 7.5% for subgenotypes, consistent with rules of classification and further validating the results obtained with the Geno2Pheno algorithm (Table 3.1).

Table 3.1 Divergence between HBV genotypes/subgenotypes present in Portugal.

	HBV_A1	HBV_A2	HBV_D3	HBV_D4	HBV_E
HBV_A1		0.041	0.091	0.086	0.086
HBV_A2	**0.006**		0.097	0.088	0.096
HBV_D3	**0.011**	**0.011**		0.049	0.082
HBV_D4	**0.011**	**0.011**	**0.007**		0.074
HBV_E	**0.010**	**0.011**	**0.010**	**0.010**	

The number of base substitutions per site from averaging over all sequence pairs between groups are shown above the diagonal and standard error estimate(s) are shown below the diagonal in bold

Genetic divergence within genotypes/subgenotypes was higher for HBV/A1 and HBV/D3 and lower for HBV/E (Table 3.2).

Table 3.2 Divergence within HBV genotypes/subgenotypes isolates present in Portugal.

Genotype/Subgenotype	Genetic distance	Standard Error
A1	0.023	0.003
A2	0.018	0.002
D3	0.022	0.003
D4	0.018	0.002
E	0.017	0.002

The number of base substitutions per site from averaging over all sequence pairs within groups are shown in the central column and standard error estimate(s) are shown in the right column.

3.1.2 HBV was first introduced in Portugal around 1857 through genotype D

The estimated mean substitution rate obtained in TRACER's analysis was $9.45\text{x}10^{-5}$ s/s/y (95% HPD: $4.22\text{x}10^{-5}$ - $1.43\text{x}10^{-4}$). Data from spatiotemporal reconstruction analyses using this substitution rate suggest that HBV might have originated in France (PSP = 1) around the year 886 (95% HPD: 414 B.C - 1514) (Table 3.3).

Among the studied genotypes, HBV/D was probably the first genotype to emerge in the beginning of XVI century, immediately followed by HBV/A. HBV/E was the last to emerge, maybe three centuries later than A (Table 3.3).

In Portugal, HBV/D4 was the first to emerge possibly due to a single introduction estimated to have occurred around 1857 (95% HPD: 1699 – 1931). HBV/D3 and HBV/A2 were introduced around 1882 (95% HPD: 1756 – 1933) and 1883 (95% HPD: 1767 – 1940), respectively. HBV/D3 was introduced again in Portugal at the end of XIX century and HBV/A2 was again introduced in Portugal twice during the first half of XX century. HBV/A1 was introduced multiple times in Portugal, the first one having occurred near 1909 (95% HPD: 1820 – 1954) and the last three until 1954 (95% HPD: 1894 – 2002). Finally, the first and second introductions of HBV/E in Portugal, occurred six years apart in 1906 (95% HPD: 1811 – 1951) and 1912 (95% HPD: 1835 – 1948) (Table 3.3).

Table 3.3 Estimated time and place of HBV origin, origin of the Portuguese strains, time of introduction in Portugal and role of Portugal in the global HBV spread.

Event	Median year [HPD 95%]	PSP	Country of origin	% of PT clustered sequences (N)		Spread from Portugal to
HBV origin	886 [414 B.C - 1514]	1	France			
GT A emergence	1537 [1037 - 1787]	1	France			
GT A arrival in Portugal	1933 [1864 - 1967]	0.9	Mali	3.03% (1)		N.O.
	1950 [1887 - 1981]	0.36	Cameroon	3.03% (1)		N.O.
	1932 [1851 - 1975]	0.98	Spain	3.03% (1)		Cuba
A1 emergence	1711 [1416 - 1855]	0.98	Libya			
A1 arrival in Portugal	1913 [1834 - 1952]	0.17	Brazil	78.8%	42.3% (11)	France, Italy, Netherlands and Cape Verde
	1909 [1820 - 1954]	0.29	Brazil		57.7% (15)	France, Italy and Germany
	1954 [1894 - 2002]	0.95	Zimbabwe	6.06% (2)		N.O.
	1923 [1846 - 1963]	0.72	Algeria	6.06% (2)		India
A2 emergence	1785 [1559 - 1891]	1	Belgium			
A2 arrival in Portugal	1883 [1767 - 1940]	1	Belgium	85.7%	38.9% (7)	Spain, Germany and France and Cape Verde
	1925 [1913 - 1936]	0.0	Belgium		61.1% (11)	Cuba and Cameron
	1944 [1885 - 1974]	0.91	Spain	14.3% (3)		N.O
GT D emergence	1501 [922 - 1790]	0.99	France			
D3 emergence	1779 [1547 - 1897]	1	France			
D3 arrival in Portugal	1882 [1756 - 1933]	0.01	France	40% (6)		Brazil, USA and Spain
	1899 [1811 - 1944]	0.5	France	60% (9)		Spain, Brazil and back to France
D4 emergence	1766 [1523 - 1885]	1	France			
D4 arrival in Portugal	1857 [1699 - 1931]	0.47	France	100%		Brazil, Germany and back to France
GT E emergence	1818 [1619 - 1914]	1	France			
GT E arrival in Portugal	1912 [1835 - 1948]	0.16	Spain	50% (16)		USA
	1906 [1811 - 1951]	0.46	Spain	50% (16)		Cape Verde and Norway

N.O, no occurrence; PSP, posterior state probability; GT, genotype; PT seq, Portuguese sequences; B.C., before Christ (calendar years not labelled with B.C. should be considered After Christ). This table summarizes the data estimated in the phylogeography analysis using a Bayesian MCMC method implemented in the BEAST v1.10.4, with the GTR+F+R4 nucleotide substitution model, an uncorrelated lognormal relaxed molecular clock model under a Bayesian skyline plot, as coalescent priors.

3.1.3 Most HBV genotypes were imported to Portugal from other European countries

HBV seems to have evolved in its original region, France, during several years, diverging into one branch that gave rise to the present genotype A, and another one that was at the origin of the remaining genotypes studied (Figure 3.2).
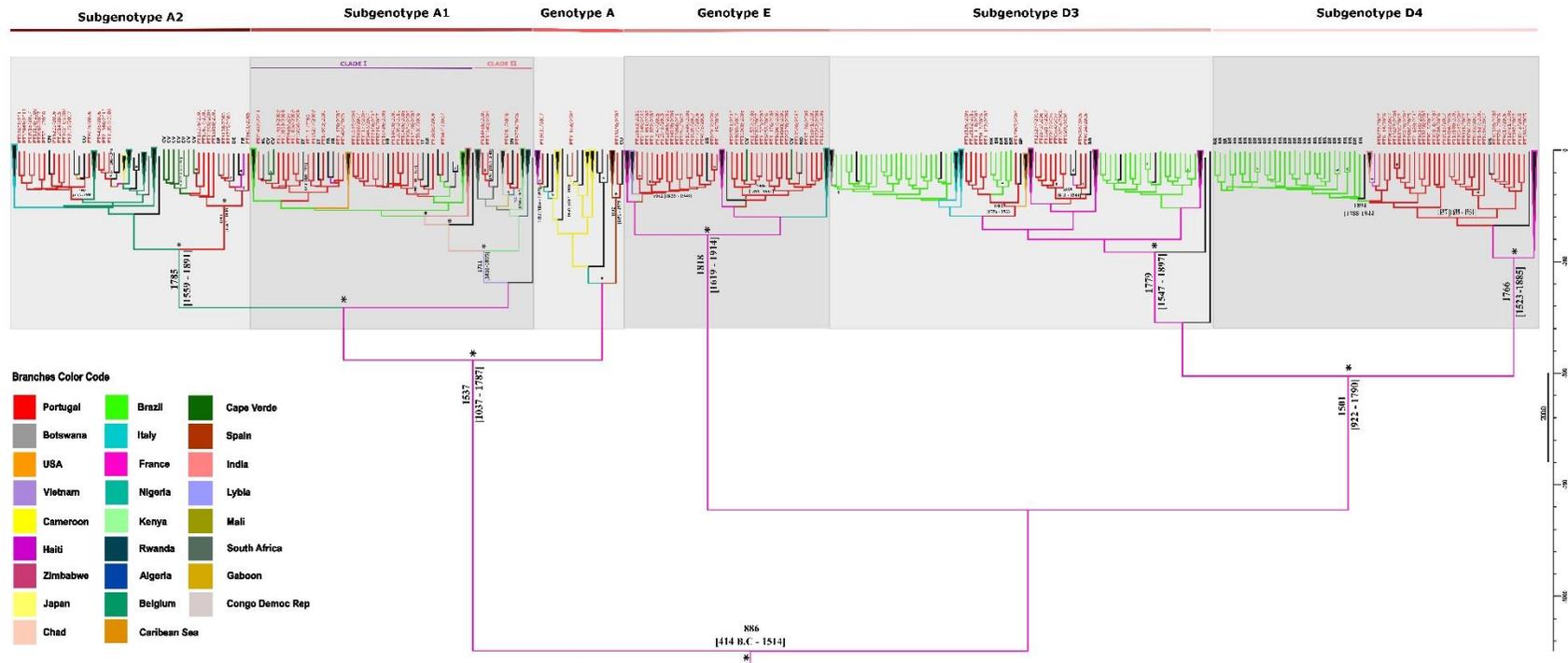
Figure 3.2 | Bayesian maximum clade credibility tree of HBV partial polymerase gene sequences. The colored internal branches represent the locations of the parental nodes, as assigned by continuous Bayesian phylogeography (color code below the tree). Branches and tips of Portuguese isolates are labeled in red. External branches labeled with black initials represent the countries to where Portugal exported HBV strains (BR-Brazil; CM-Cameroon; CU-Cuba; CV-Cape Verde; DE-Germany; FR-France; IN-India; IT-Italy; NL-Netherlands; NO-Norway; SP-Spain; US-United States of America). The external branches representing Portuguese speaking countries, Brazil, and Cape Verde, are highlighted in light green and dark green, respectively. Collapsed branches (triangles) represent clades that are not of primary interest for the study and are too large for displaying in full. They are filled in black because they include isolates from different origins, being outlined with the color of the country that originated them. Within subgenotype A1, Clade I – A1 Asian/American clade is labeled with a purple line, while Clade II – A1 African Clade is labeled with a rose line. The scale on the right of the tree represents years before the sampling time of the most recent sequence (2012). Tree is annotated with genotypes/subgenotypes emergence dates and dates of introduction in Portugal. Branches with a posterior state probability support ≥0.9 are annotated with *. More information about reference sequences (origin, date of collection and accession number) is available in annex 2_Table A2. The complete MCC tree without collapsed branches is available in annex 4_Figure A4.

In Portugal, HBV/D3 and HBV/D4 were both introduced via France. The emergence of HBV/A2 in Portugal occurred via Belgium and Spain, while HBV/E was exclusively introduced in Portugal via Spain. The unique subgenotype that entered in Portugal via non-European countries was HBV/A1 (Figure 3.2; Table 3.3).

Most of the HBV/A1 Portuguese strains cluster in A1 clade I (purple line) known as the Asian-American Clade, and only four Portuguese strains cluster together with the A1 clade II (rose line) known as the African Clade (Figure 3.2).

3.1.4 Portugal exported HBV/A1 and HBV/A2 to several European countries and Cape Verde

Our results indicate that HBV/A diversified in North Africa around 1711 (95% HPD: 1416 - 1855) originating HBV/A1, and in Belgium near 1785 (95% HPD: 1559 - 1891) originating HBV/A2 (Figure 3.2; Table 3.3). Possibly, HBV/A1 had its origin in the region of Libya and migrated across countries such as Chad. From this region, it seems to have been introduced in Brazil and India near 1869 (95% HPD: 1731 – 1933). Finally, coming from Brazil, it was introduced in Portugal more than once around 1894 (95% HPD 1792 – 1942) and 1904 (95% HPD 1819 – 1945), later spreading from Portugal to other countries in Europe, mainly Italy and France, as well to Cape Verde Islands in Africa. A second route of A1 subgenotype introduction in Portugal occurred directly via Africa, possibly through Algeria and Zimbabwe around 1923 (95% HPD 1846 - 1963) and 1954 (95% HPD 1894 - 2002), respectively (Figure 3.3).
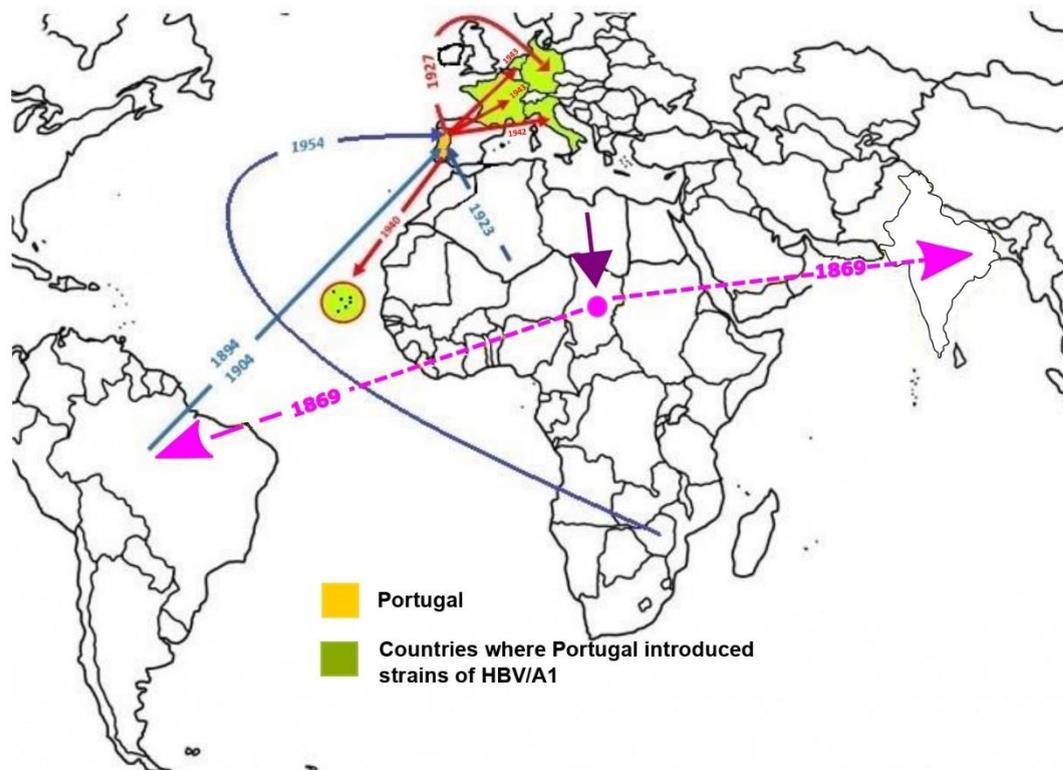
Figure 3.3 | Probable major dispersal pathways of HBV/A1 as estimated by phylogeographic analysis. The blue arrows represent probable routes of introduction of the HBV/A1 in Portugal. The red arrows represent the probable dispersion paths of HBV/A1 from Portugal towards other countries. A significant dispersion event, prior to the introduction of A1 in Portugal, is represented in purple and pink: the purple arrow is located on the probable country of emergence of HBV/A1 indicating its dissemination direction; the pink circle represents the region from which HBV/A1 has been exported outside Africa. The dashed pink arrows simulate the likely route of exportation, that seemingly occurred simultaneously to Brazil and India.

In a different pathway, HBV/A2 has possibly emerged in Belgium between the middle of XVI and the end of XIX century, after migration from its original source in France (Figure 3.2; Table 3.3). After introduction from Belgium, HBV/A2 was established in Portugal in 1883 (95% HPD 1767 – 1940) and in 1925 (95% HPD 1913 - 1936) and spread not only to Spain, Germany, and France but also to Cuba and African countries such as Cameroon and Cape Verde. An additional HBV/A2 introduction in Portugal occurred around 1944 (95% HPD 1885 – 1974) via Spain (Figure 3.4).
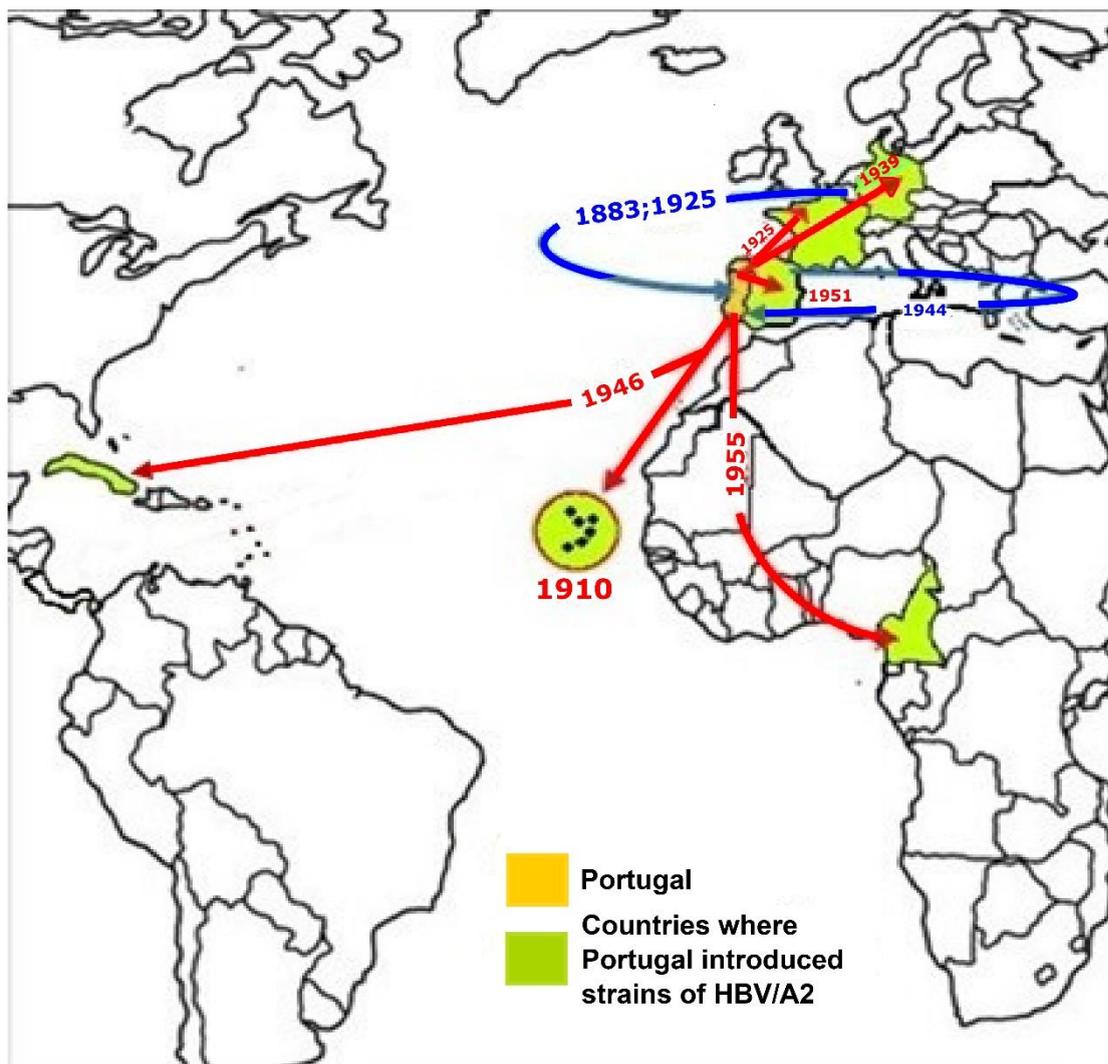
Figure 3.4 | Probable major dispersal pathways of HBV/A2 as estimated by phylogeographic analysis.
The blue arrows represent probable routes of introduction of the HBV/A2 in Portugal. The red arrows represent the probable dispersion paths of the HBV/A2 from Portugal towards other countries.

Our analysis revealed that the three Portuguese strains to which no subgenotype A classification was assigned have different origins. While one of them entered Portugal around 1932 (95% HPD: 1851 – 1975) via France and then Spain, the other two strains had an African origin. After divergence from the common ancestor, these strains seem to have diverged further in Nigeria, Cameroon, and Mali before finally being imported to Portugal. These last two strains were introduced in Portugal around 1933 (95% HPD: 1864 – 1967) and 1950 (95% HPD: 1887 – 1981) (Figure 3.2; Table 3.3).

3.1.5 Portugal exported HBV/D4 to Brazil around 1894, causing a local epidemic

The common ancestor of D subgenotypes was dated to diverged in France around 1501 (95% HPD: 922 – 1790), branching off into subgenotypes D3 and D4 likely in the second half of XVIII century (Figure 3.2; Table 3.3). The introduction of HBV/D3 in Portugal occurred via France in two different occasions. It entered Portugal around the year of 1882 (95% HPD: 1756 - 1933) and then was exported to Brazil near 1909 (95% HPD: 1814 – 1952), while also spreading to USA and Spain (Figure 3.5_D3). The second importation to Portugal occurred near 1899 (95% HPD 1811 - 1944). These isolates were exported once again to Brazil and Spain. From this last country, they were later introduced again in Portugal from where, via France, they spread not only to Germany but also to Cuba (Figure 3.2). Other countries exporting HBV D3 to Brazil were France and Italy (Figure 3.2; Figure 3.5_D3).

Subgenotype D4 originated in France, possibly entered Portugal near 1857 (95% HPD 1699 - 1931) from where it seems to have been exported to Brazil around 1894 (95% HPD 1788 – 1944), generating a cluster of Brazilian HBV/D4 strains. Portugal also exported HBV/D4 to France and Germany (Figure 3.2; Figure 3.5_D4).
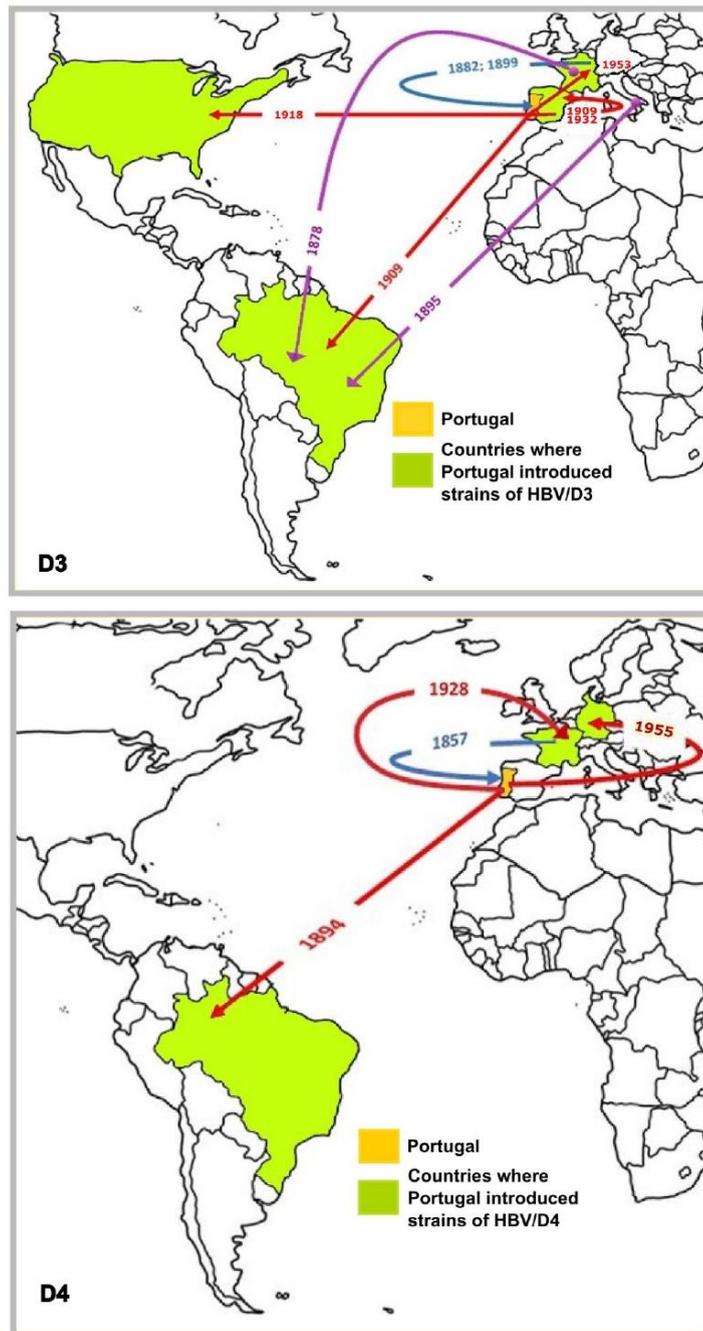
Figure 3.5 | Probable major dispersal pathways of subgenotype HBV/D3 (above) and HBV/D4 (below) as estimated by phylogeographic analysis. The blue arrows represent probable routes of introduction of the subgenotypes in Portugal. The red arrows represent the probable dispersion paths of subgenotype from Portugal towards other countries. Purple arrows show the contribution of other countries, besides Portugal, to the introduction of D3 subgenotype in Brazil.

3.1.6 Genotype E was exported from Portugal to Cape Verde around 1940

Our data indicates that genotype E probably emerged more recently than all the others near 1818 (95% HPD: 1619 - 1914) in France (Figure 3.2; Table 3.3). After migrating to Spain, it was introduced in Portugal around 1906 (95% HPD: 1811 - 1951) and 1912 (95% HPD 1835 - 1948) creating two subclusters of Portuguese sequences (Figures 3.2, Figure 3.6).
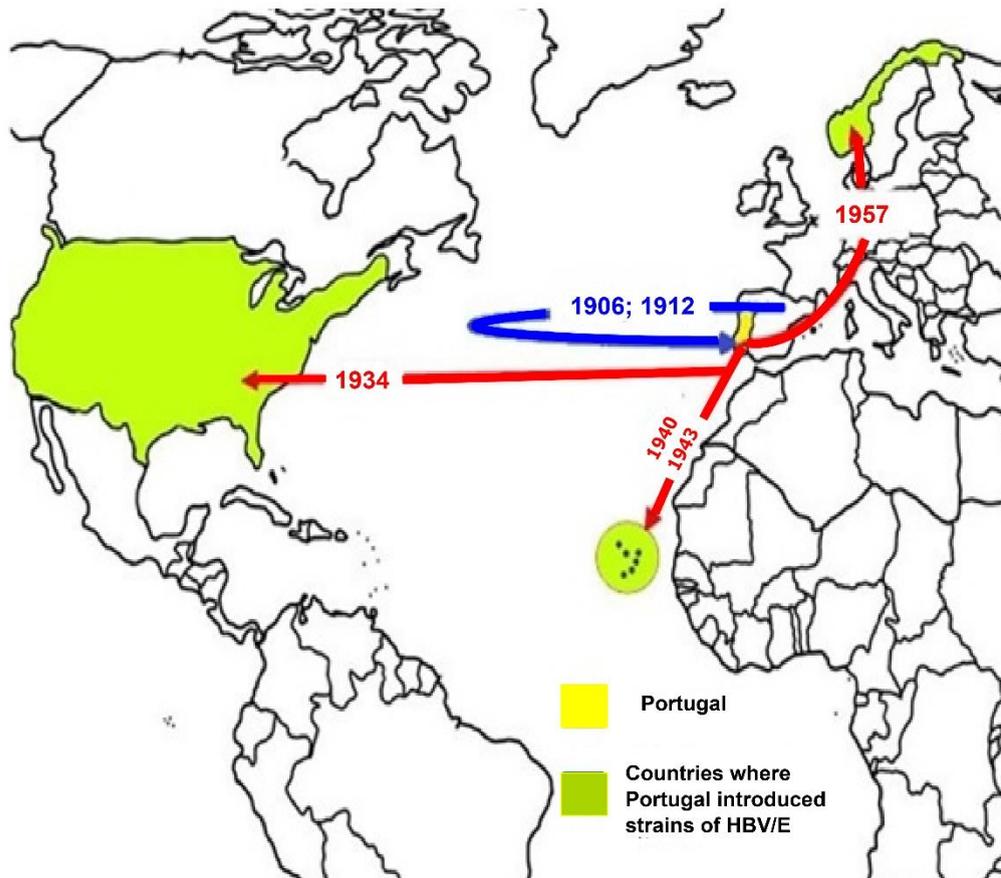
Figure 3.6 | Probable major dispersal pathways of HBV/E as estimated by phylogeographic analysis. The blue arrows represent probable routes of introduction HBV/E in Portugal. The red arrows represent the probable dispersion paths of HBV/E from Portugal towards other countries.

One of the clusters seems to be responsible for the exportation of HBV/E to Cape Verde and Norway region, while the other seems related to HBV/E spread to USA (Figures 3.2; Figure 3.6; Table 3.3).

3.2 Epidemic history of HCV genotypes and subtypes in Portugal

3.2.1 HCV genotypes and subtypes circulating in Portugal

A total of 230 patients were included in this study. Overall, 59.1% (n=136) of subjects were men and had a median age of 41 years [Interquartile Range (IQR): 49 – 36)]. The genotyping results showed a majority of HCV/1 (62.6%; n=144), followed by HCV/3 (18.3%; n=42) and HCV/4 (15.9%; n=37), while HCV/2 was the least frequent (3.0%; n=7) (Table 3.4). Among patients harboring HCV/1, the most frequent subtype was 1a (75.1%, n=108) followed by HCV/1b with a 3.1-fold lower prevalence (24.4 %, n=35), (P <0.0001); only one patient was infected with HCV subtype 1g (HCV/1g) (0.4%). Among patients with HCV/2, 1.3% (n=3) of the total population had an undetermined subtype, while subtype 2a (HCV/2a) and subtype 2C (HCV/2c) were each found in 0.9% (n=2) of the total population. All HCV/3 were subtype 3a (HCV/3a). Among patients with HCV/4, the most represented subtypes were 4a (HCV/4a), with 10.4% of the total population (n=24) and 4d (HCV/4d) (4.3%; n=10); Subtype 4b (HCV/4b), 4f (HCV/4f) and 4k (HCV/4k) were found in one (0.4%) patient each (Table 3.4, Fig. 3.7).

Table 3.4 Relative frequencies of genotypes, subtypes, and HCV/1a clades I and II.

HCV genotype/subtype	Age, median (IQR)[#]
N = 230 (%)	
1a clade I	23 (10.0)	37 (33–40)
1a clade II	85 (37.0)	40 (36–47)
1b	35 (15.2)	53 (36–60)
1g	1 (0.4)	53
2	3 (1.3)	57
2a	2 (0.9)	44
2c	2 (0.9)	47
3a	42 (18.3)	42 (35–48)
4a	24 (10.4)	40 (34–46)
4b	1 (0.4)	69
4d	10 (4.3)	42 (39–46)
4f	1 (0.4)	60
4k	1 (0.4)	46

 # IQR, interquartile range (197)

The HCV/1a clade II was significantly more prevalent than clade I (78.7%; n=85/108 vs. 21.3%; n=23/108; P<0.0001). Interestingly, when the HCV/1a lineages assigned by

phylogenetic analysis were analyzed with the Geno2Pheno [HCV] algorithm, 25 sequences had inconsistent clade results (Annex 5_Table A5). Likewise, inconsistent results were observed for two sequences that were classified as HCV/2 and HCV/4b by phylogenetic analysis and as HCV/2b and HCV/4w by Geno2Pheno [HCV].

The various polyphyletic patterns observed in the phylogenetic tree suggested multiple and old introductions of the different HCV subtypes in this population (Figure 3.7).

Figure 3.7 | Phylogenetic analysis of NS5B gene sequences from HCV infected patients attending the Hospital Santa Maria, Lisbon, Portugal. The sequences from Portuguese patients are colored in red and reference HCV sequences are colored in black. HCV genotypes and subtypes are indicated with different colored strips with subtype 6 as an outgroup. The bootstrap values supporting the internal branches defining a genotype, subtype or clade are shown at the nodes as blue-filled circles with size corresponding to the magnitude of bootstrap values (only values between 70 and 100% are shown). Bootstrap values of 70% or greater provide reasonable confidence for assignment of an individual sequence to one or the other genotype. The scale represents number of base substitutions per site (197).

3.2.2 Origin and epidemiologic history of HCV strains circulating in Portugal

The dates of the MRCA of the various HCV subtypes circulating in the Portuguese population were estimated in BEAST under an ULCD with a BSP coalescent demographic model, using a normal prior distribution for the UCLD mean clock rate parameter (0.001 ± 0.0001 s/s/y) (Table 3.5).

Table 3.5 Estimated dates of MRCAs for HCV subtypes identified in the current study.

Dataset	Dates of MRCA* (95% HPD interval)
Subtype 1a (n = 108)	1950 (1922, 1973)
Subtype 1b (n = 35)	1946 (1847, 1976)
Subtype 3a (n = 42)	1963 (1947, 1977)
Subtype 4a (n = 24)	1988 (1980, 1995)
Subtype 4d (n = 24)**	1982 (1964, 1995)

*Mean estimates of MRCA dates in calendar years.
**Contains additional HCV/4d sequences (n = 14) from a previous study in Portugal (197).

According to our estimates, the ancestor of the HCV/1a strains circulating in Portugal dated back to 1950 (95% HPD: 1922 - 1973) whereas that of HCV/1b dated back to 1946 (95% HPD: 1847 - 1965) and HCV/3a to 1963 (95% HPD: 1947 - 1977). For genotype 4 strains, similar ancestor dates were obtained [HCV/4a, 1988 (95% HPD: 1980, 1995); HCV/4d, 1982 (95% HPD: 1964, 1995)]. To investigate the epidemiologic history of the different HCV subtypes in the Portuguese population, BSP reconstruction was made for each subtype (Figure 3.8).
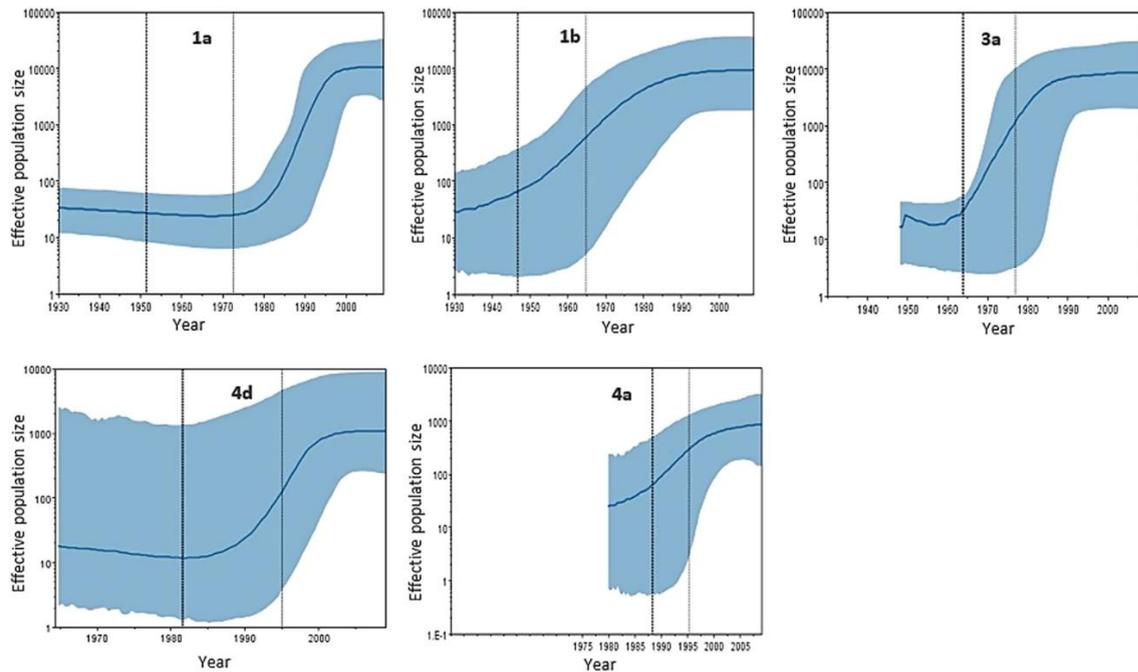
Figure 3.8 | Epidemic history of HCV subtypes in Lisbon, Portugal. Bayesian skyline plot showing the epidemic history of the most prevalent HCV subtypes (1a, 1b, 3a, 4a and 4d) found in patients attending the Hospital Santa Maria of Lisbon, Portugal. The solid line represents the changes in the mean effective population size (Ne) through time on a $\log_{10}$ scale, with the shaded area corresponding to the 95% HPD interval. The bold dotted and faint dashed black vertical lines represent the median and upper boundaries of the time to the MRCA, respectively. For HCV/4d, plots were built using additional sequences from Portugal (n = 14) retrieved from GenBank (197).

The data indicates that the first significant increase in HCV prevalence in Portugal occurred in the first half of the XX century (the 1930s) and was caused almost exclusively by HCV/1b until the 1960s. At this time, a second epidemic emerged caused by HCV/3a; a third epidemic caused by HCV/1a emerged during the 1980s. The first two epidemics caused by HCV/1b and HCV/3a reached equilibrium at the end of the 1980s, whereas the HCV/1a epidemics continued to grow rapidly until the end of the 1990s when it reached a plateau. The most recent entries of HCV in Portugal were caused by HCV/4a and HCV/4d and while the HCV/4d epidemics seems to have reached a plateau, HCV/4a epidemics is still expanding although with a lower growth rate when compared to the exponential growth rate phase of HCV/1a and HCV/3a (Figure 3.8).

3.3 Evolutionary history and phylogeography of Clades I and II of HCV subtype 1a in Portugal

Given that HCV/1a is the most frequent in Portugal, with two different clades circulating among these infected individuals, and with clade II unusually more prevalent than in other regions of Europe, as seen in the previous section 2, the evolutionary history of both clades was investigated.

3.3.1 Phylogenetic analysis of the source population

Phylogenetic analysis of the source group of 83 samples, showed that twenty-five samples were classified as HCV/1b (30.1%) and 58 classified as HCV/1a (69.9%) (Figure 3.9).
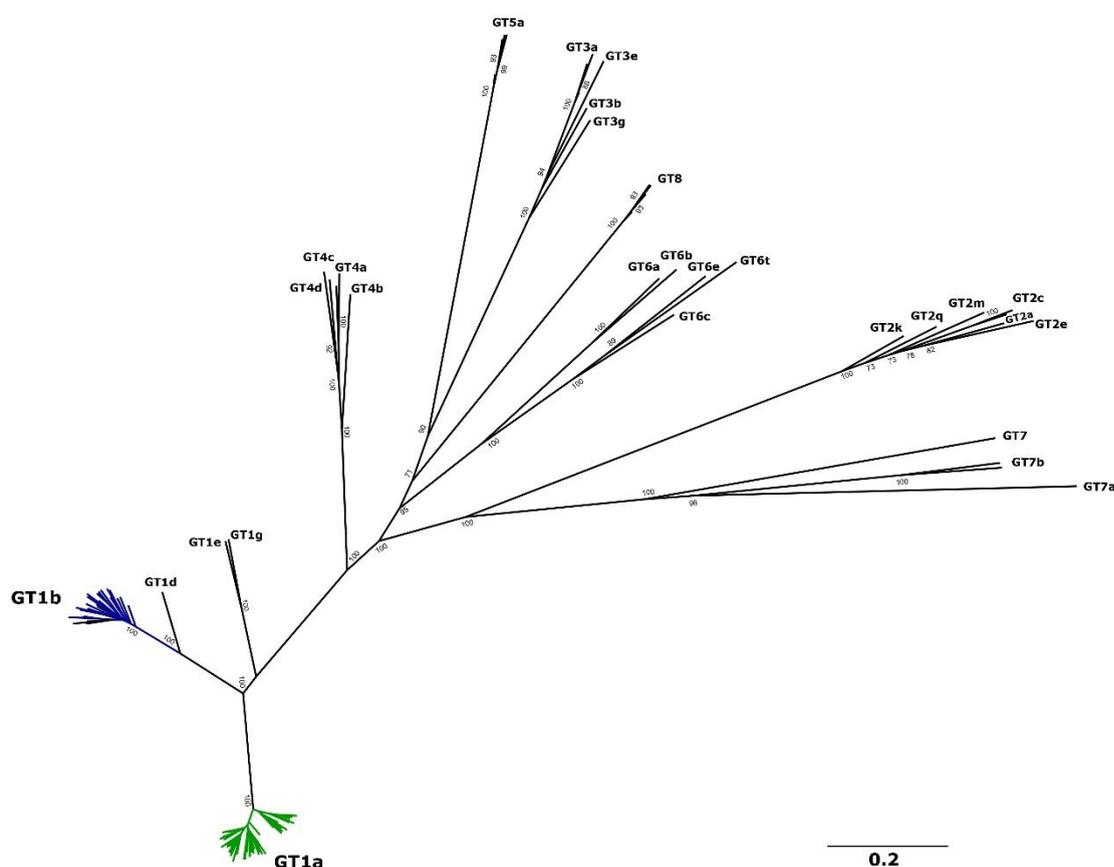
Figure 3.9 | ML phylogenetic tree reconstruction showing the classification of 83 Portuguese HCV samples into HCV/1b and HCV/1a. The tree was reconstructed using IQtree 1.6.1 software under a TIM2+F+I+G4 evolutionary model, with 1000 bootstraps replicates Subtype1b is represented by the blue external branches and subtype 1a by the green external branches. Genotypes 2, 3, 4, 5, 6, 7 and 8 are represented along the tree, each one showing some examples of their subtypes. Bootstraps above 70% are annotated in the branches. The scale in the bottom right represents the number of substitutions per site.

3.3.2 Characteristics of the selected population

A total of 58 DAA-naïve patients with chronic HCV subtype 1a infection were enrolled in the study with a mean age of 47 years old and 81% of the patients belonging to male gender (table 3.6). Among the patients 58.6% (n=34) were coinfected with HIV-1 and only around 25.9% (n=15) of the patients showed a single HCV infection. For the remaining patients, 13.8% (n=8) and 1.7% (n=1) were HCV/HIV/HBV and HCV/HBV coinfected, respectively. HCV Log_{10} viral load mean was 6.25 [(Standard Deviation (SD): +/- 0.54)].

Among the 58 patients included in this study, 15 (25.9%) and 43 (74.1%) were infected with HCV 1a clade I and HCV 1a clade II, respectively. It seems there are more males infected with clade II than with clade I and more females infected with clade I than clade II. Statistical analysis revealed that there is an association between gender and HCV/1a clade of infection (p=0.025) (table 3.6).

Table 3.6 Demographical and clinical data stratified by clade I and clade II.

Population Data	All patients	Clade I	Clade II	P
Demographical Data				
N, (%)	58	15 (25.9%)	43 (74.1%)	—
Age, years [Mean (SD)]	47 (7.08)	45 (5.81)	48 (7.38)	0.170
Gender				
Male (n, %)	47 (81.0%)	9 (60.0%)	38 (88.4%)	**0.025**
Females (n,%)	11 (19.0%)	6 (40.0%)	5 (11.6%)	
Viral Load (Log$_{10}$)				
Mean	6.25	6.27	6.24	0.829
SD	0.54	0.32	0.59	
Co-infection/Monoinfection				
HCV	15 (25.9%)	5 (33.3%)	10 (23.3%)	
HCV/HIV (n, %)	34 (58.6%)	8 (53.3%)	26 (60.5%)	0.502
HCV/HBV (n, %)	1 (1.7%)	N.O.	1 (2.3%)	
HCV/HIV/HBV (n, %)	8 (13.8%)	2 (13.3%)	6 (13.9%)	

Mean age, viral load log_{10} and co-infection/monoinfection status did not differ significantly between both clades (p>0.05).

3.3.3 Clade 2 sequences are dominant in Portugal and associated with males

Phylogenetic analysis showed that among the 58 Portuguese HCV/1a sequences, 74.1% (n=43) belong to clade II and 25.9% (n=15) belong to clade I (Figure 3.10). These results were consistent with the classification of Geno2Pheno [HCV] 0.92 (Annex 6_Table A6). VERSANT® HCV Genotype 2.0 LiPA show genotypes and subtypes results but do not classify HCV/1a samples into clades. Nonetheless, as for genotyping/subtyping classification two results disagreed with phylogenetic analysis and Geno2Pheno. One of them was due to inconclusive LiPA genotyping and the other due to a wrong genotype attribution by LiPA (Annex 6_Table A6).

Clade I isolates present a segregation in two distinct monophyletic subclades, named here Y and X, with 60.0% and 40.0% of the Portuguese sequences belonging to each one, respectively.
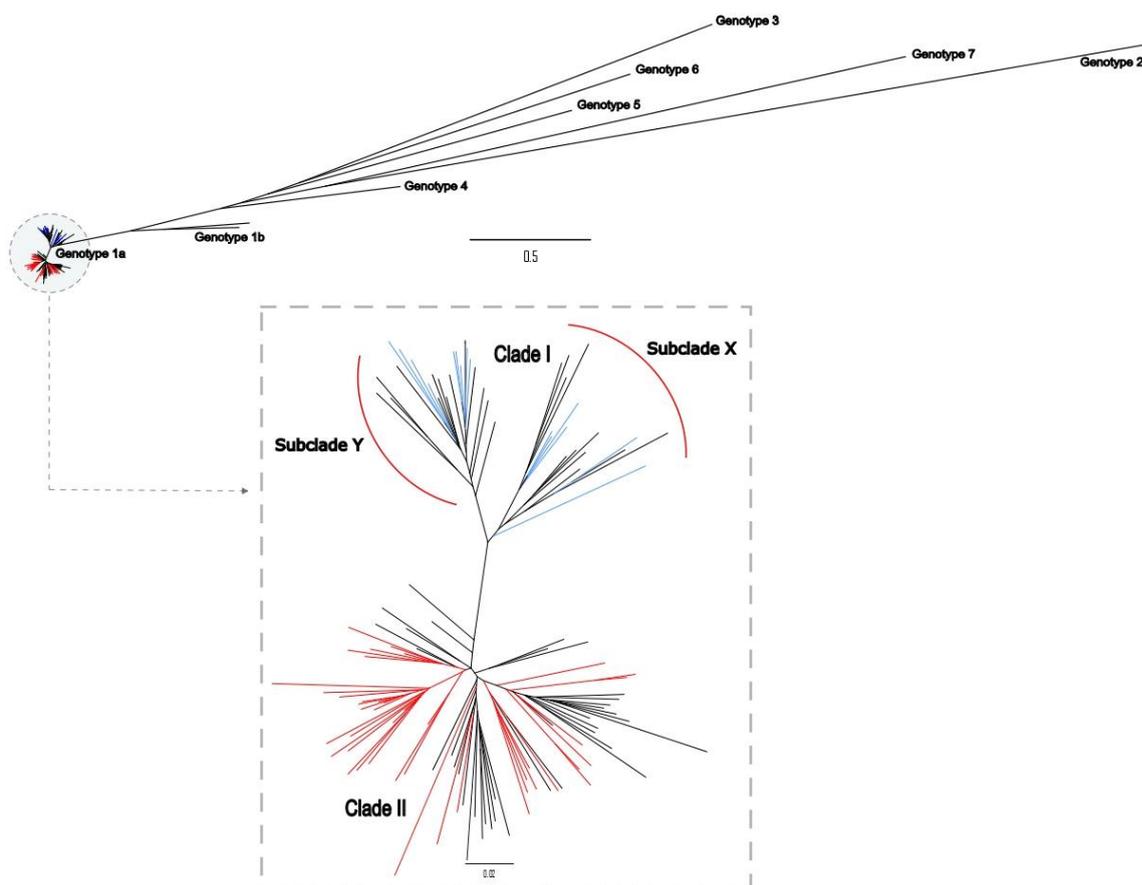
Figure 3.10 | ML phylogenetic tree of HCV NS5A sequences showing evolutionary relationships between HCV strains. The unrooted Maximum Likelihood phylogenetic tree shows that Portuguese strains belong to subtype 1a. The inset shows in more detail which sequence belong to clades I (blue lines) and II (red lines). The red round lines show two subclades of clade I.

3.3.4 Origin and dispersal routes of clades I and II

The analysis in TempEst software showed a temporal structure present in both datasets, as revealed by the correlation coefficient between genetic divergence and time (0.3887 for clade I and 0.2544 for clade II).

The Clade I seems to have had its origin in USA (Figure 3.11), while clade II origin is linked to France (Figure 3.12).

Figure 3.11 | Bayesian MCC tree of HCV *NS5A* gene sequences from clade I. Internal branches are colored according to highest posterior state probability of the parental nodes (i.e., the most likely location of the ancestral nodes), showing HCV/1a clade I geographical dispersion from its origin until it arrives in Portugal (color code below the tree). External branches are labeled with the initials of the countries of sampling of each sequence (additional information on country initials codes in annex 7). Branches and initials of Portugal are shown in red, and the remaining-colored external branches represent countries to where Portugal exported clade I strains. The scale at the bottom right of the tree represents years before the sampling time of the most recent sequence (2017).

According to our analyses, both clades emerged in 1934 with a 95% HPD interval between 1918 and 1947 (Table 3.7).
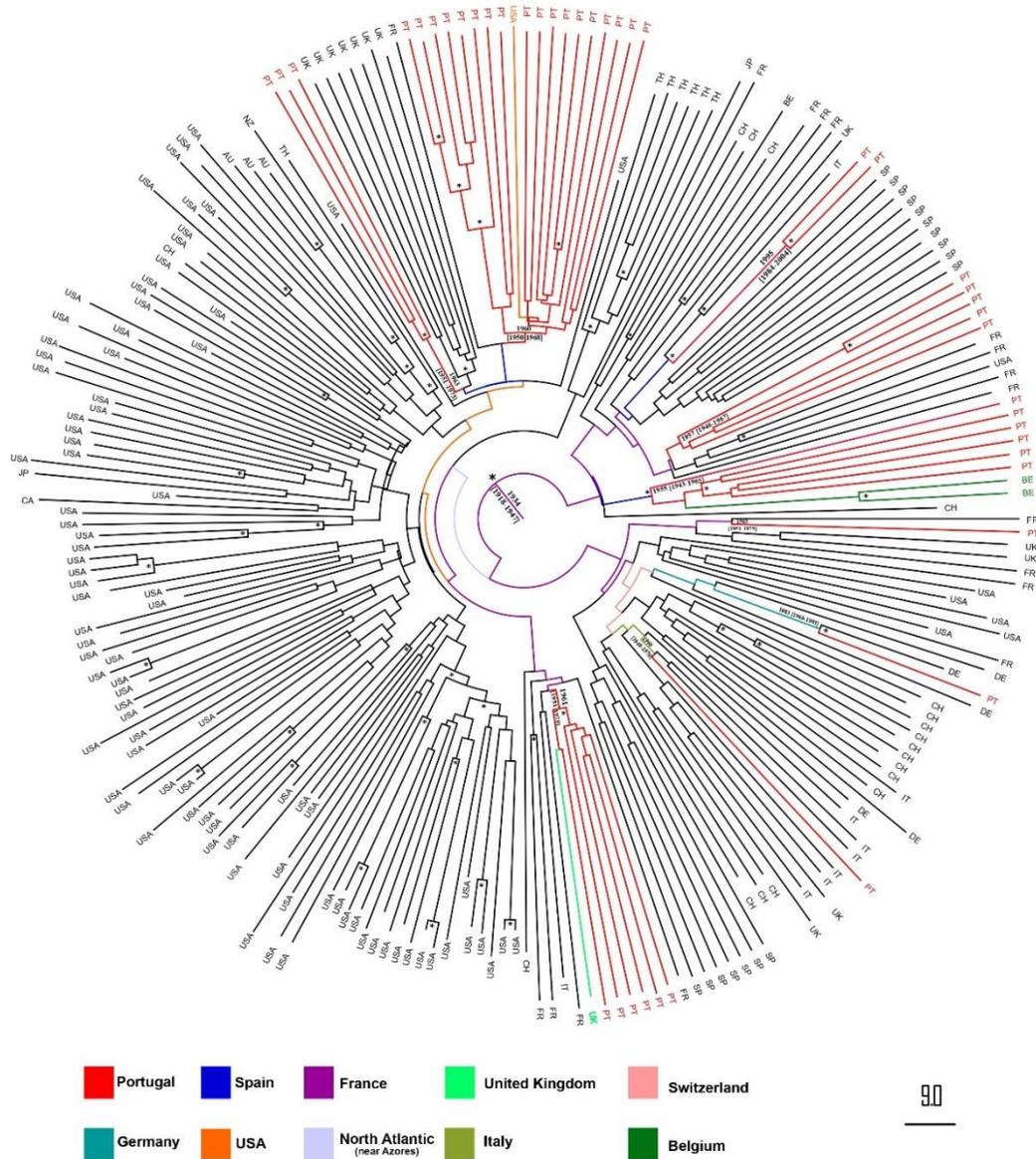
Figure 3.12 | Bayesian MCC tree of HCV *NS5A* gene sequences from clade II. Internal branches are colored according to highest posterior state probability of the parental nodes (i.e., the most likely location of the ancestral nodes), showing the geographical dispersion of HCV/1a clade II from its origin until it arrives in Portugal (color code below the tree). External branches are labeled with the initials of the countries of sampling of each sequence (additional information on country initials codes in annex 7). Branches and initials of Portugal are shown in red, and the remaining-colored external branches represent countries to where Portugal exported clade II strains. The scale at the bottom right of the tree represents years before the sampling time of the most recent sequence (2016).

Clade I was introduced multiple times in Portugal. The first introductions were around the second half of 1960 decade coming from USA and from France (Figure 3.11; Table 3.7).

Table 3.7 Estimated time and geographical locations of HCV/1a clades I and II origin, time of introduction in Portugal and role of Portugal in HCV spread.

Event	Median year [95% HPD]	Sub clade	Country of origin	Median year of introduction in Portugal [95%HPD]	% of Portuguese clustered sequences (N)	Spread from Portugal into
HCV Clade I origin	1934 [1918-1947]		USA			
		X	USA	1967 [1958-1975]	6,7 % (1)	Spain
			USA	1965 [1957-1972]	26,7 % (4)	Spain and Germany
Clade I			France	1995 [1984-2004]	6,7 % (1)	N.O
		Y	France	1969 [1961-1978]	20% (3)	Spain
			Spain	1975 [1965-1987]	40% (6)	N.O
HCV Clade II origin	1934 [1918-1947]		France			
			France	1961 [1951-1970]	13,5% (6)	United Kingdom
			France	1963 [1951-1975]	2,3% (1)	N.O
			France	1957 [1946-1967]	11,6% (5)	N.O
			Spain	1995 [1984-2004]	4,6 (2)	N.O
Clade II			Spain	1955 [1943-1965]	13,5% (6)	Belgium
			Spain	1960 [1950-1968]	41,9% (18)	USA
			Spain	1963 [1951-1973]	7,0% (3)	N.O.
			Italy	1960 [1949-1970]	2,3% (1)	N.O
			Germany	1983 [1968-1993]	2,3% (1)	N.O

N.O, no occurrence; HPD, High Posterior Density. This table summarizes the data estimated in the phylogeography analysis using a Bayesian MCMC method implemented in BEAST v1.10., with a HKY nucleotide substitution model, a SRD06 two-codon partitioning model, an uncorrelated lognormal relaxed molecular clock model under a Bayesian skyline plot, as coalescent priors.

Later, in the middle of the 90s it was introduced again from France. However, several of the strains circulating in Portugal derived from a single introduction in 1975 from Spain. Portugal also exported clade I strains often and almost exclusively to Spain (Figure 3.11; Table 3.7). This circulation of clade I strains can be easily visualized in Figure 3.13.

The 2 subclades of clade I observed in figure 3.10 were classified as X and Y. All clade I

Portuguese sequences with origin in USA are grouped in subclade X along with only one sequence with a French origin, while those derived from Spain and France grouped in the subclade Y (Table 3.7).
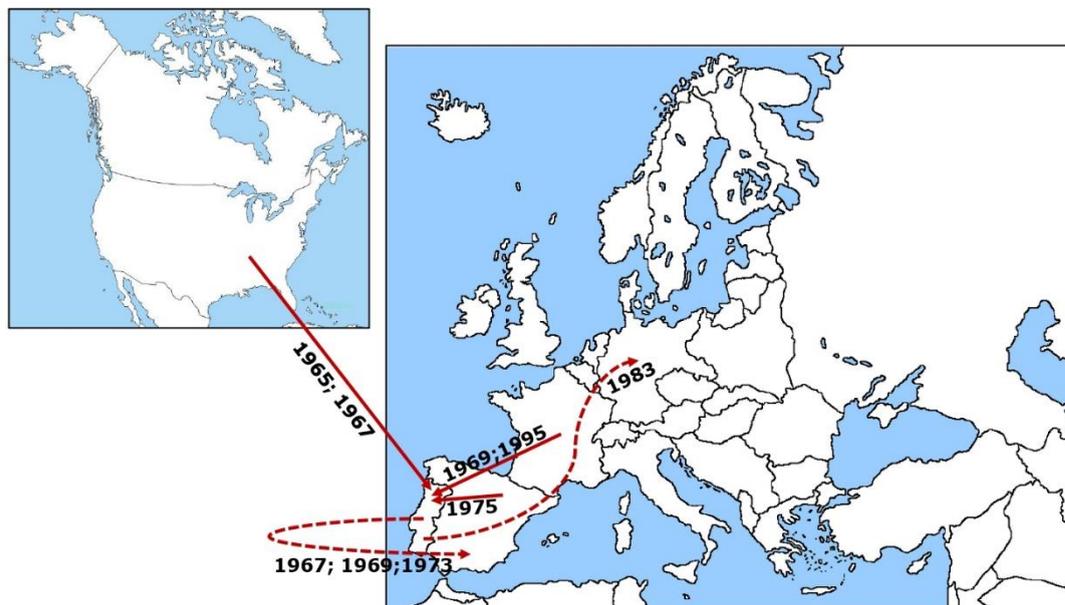
Figure 3.13 | Probable major dispersal pathways of HCV/1a clade I as estimated by phylogeographic analysis. The solid arrows represent probable routes of introduction of clade I in Portugal. The dashed arrows represent the probable dispersion paths of clade I from Portugal towards other countries. Near each arrow is represent the strains' introduction median year date.

Clade II was also introduced in Portugal multiple times. Clade II likely arrived at Portugal a decade earlier than clade I [1955 (95% HPD 1943 - 1965) vs. 1965 (95% HPD 1957 - 1972)] from Spain. Spain contributed with clade II strains several times throughout the early years of the 60s and around 1995 (Figure 3.12; Table 3.7). Between the late 1950s and the early 1960s, France was possibly the second major contributor of clade II viral strains to Portugal. Clade II strains from Italy and Germany were also introduced in Portugal, around 1960 (95% HPD 1949 - 1970) and 1983 (95% HPD 1968 - 1993), respectively (Figure 3.12; Table 3.7). Unlike clade I, our results suggest that Portuguese clade II strains did not disseminate to Spain, but were disseminated at least once to the United Kingdom, Belgium, and USA. Figure 3.14 shows this circulation of clade II strains.
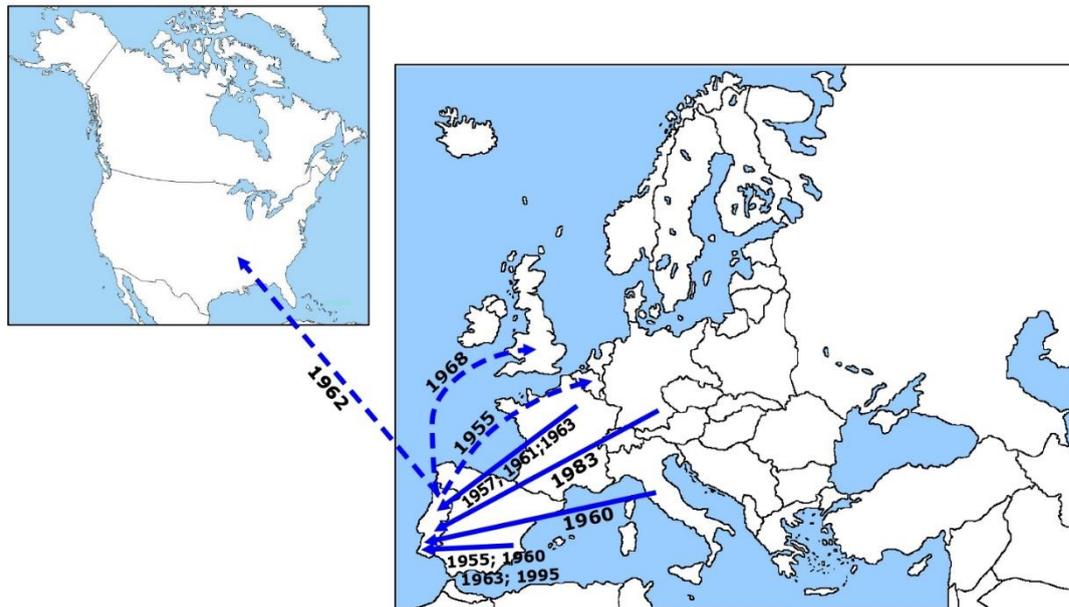
Figure 3.14 | Probable major dispersal pathways of HCV/1a clade II as estimated by phylogeographic analysis. The solid arrows represent probable routes of introduction of clade II in Portugal. The dashed arrows represent the probable dispersion paths of clade II from Portugal towards other countries. Near each arrow is represent the strains' introduction median year date.

3.3.5 HCV clade II has a wider global distribution than clade I

After the emergence around 1934, the number of clade I infections began to increase worldwide between the first years of the 50s until about the 1980s and then increased again in late 90s - early 2000s (Figure 3.15_A). It never achieved an effective population size higher than 5000. Around 2009 it declined abruptly achieving an Ne lower than 1000 in 2011 and stabilizing after that.
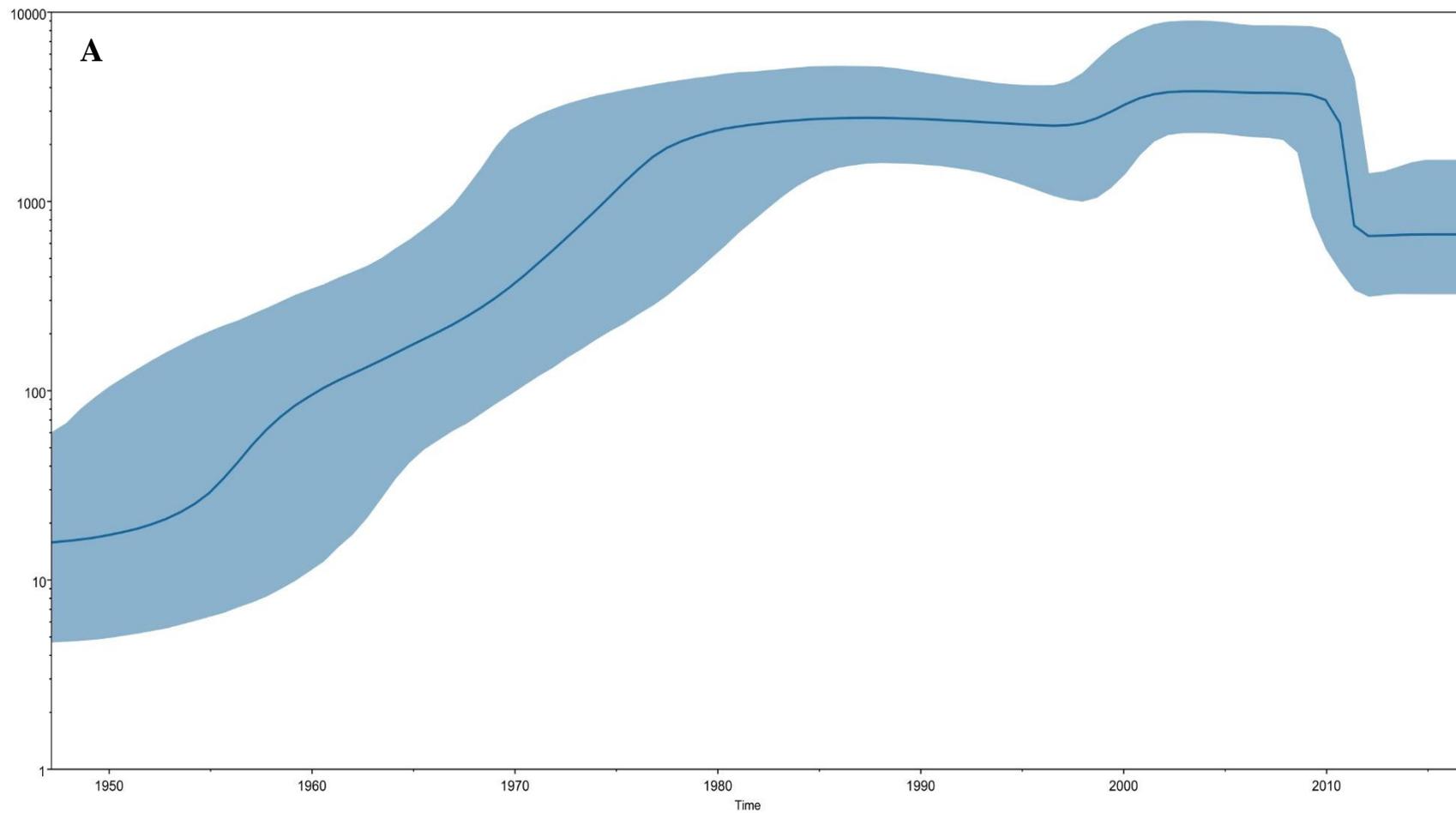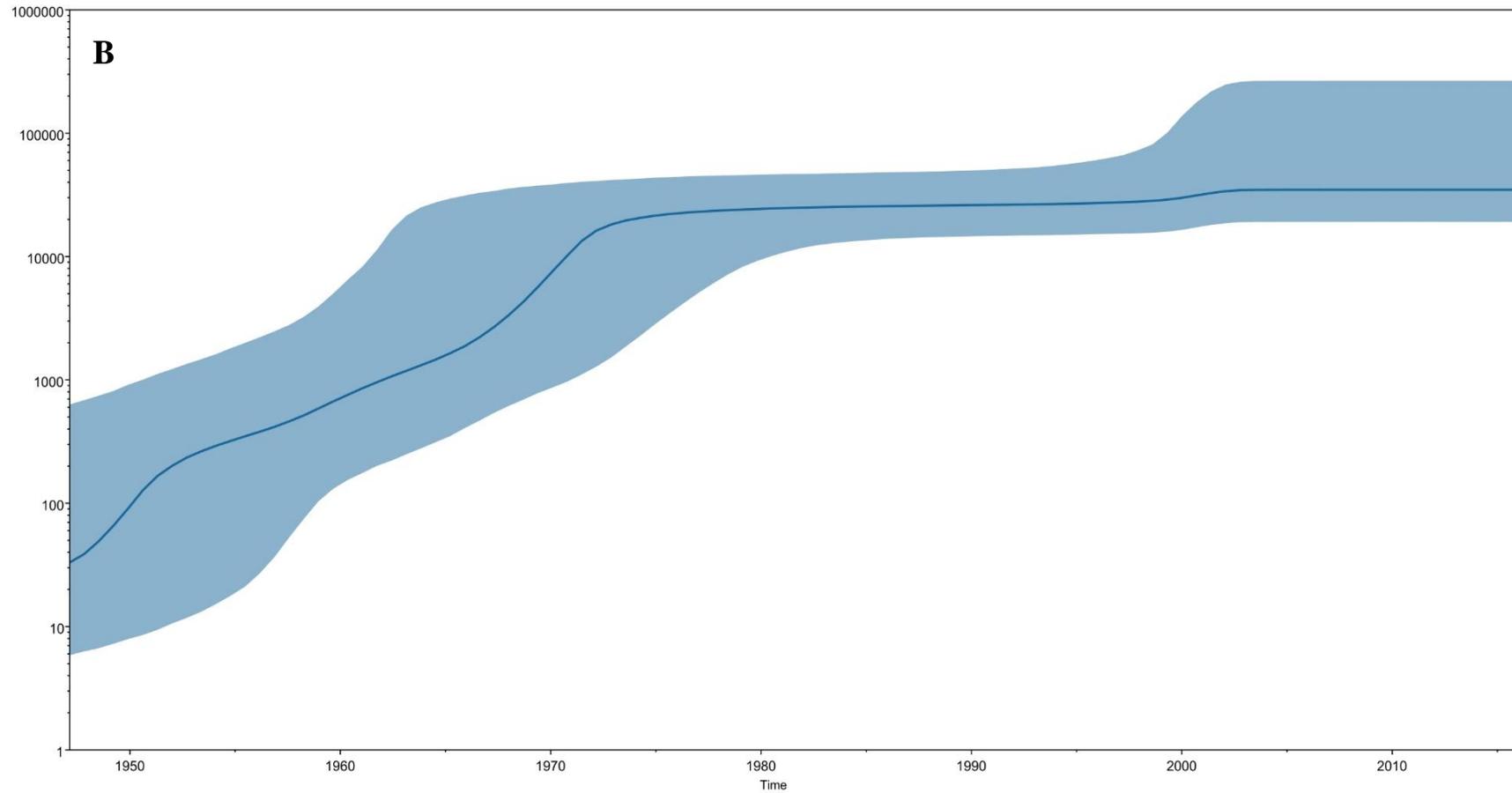
Figure 3. 15 | Bayesian Skyline Plot showing the global demographic history of HCV clade I (A) and clade II (B). The changes in the effective population size (Ne) through time is represented as a solid line and the shaded area around it shows the associated 95% HPD interval.

Clade II population growth was fast immediately after its emergence, until around 1972 (Figure 3.15_B). From then on, the epidemic stabilized with a very discrete increase in late 90s - early 2000s achieving an Ne 10-fold higher than clade I (~ 50 000).

3.4 Characterization of NS5A RAS and treatment response for the two clades of HCV/1a

3.4.1 Clade type is associated to the higher prevalence of NS5A RAS but does not affect treatment outcome

Among the 58 patients with HCV/1a infection, the relapse rate was almost twice as high in clade I (13.3%) when compared with clade II (7%) (Table 3.8). However, no statistical significance was found in this context (p = 0.587).

Table 3.8 Treatment outcome and prevalence of NS5A RASs in HCV/1a clades I and II.

Population Data	All patients	Clade I	Clade II	P
DAA Treatment Response				
SVR (n, %)	53 (91.4%)	13 (86.7%)	40 (93.0%)	0.587
Relapse (n, %)	5 (8.6%)	2 (13.3%)	3 (7.0%)	
RAS NS5A (n, %)	15 (25.9%)	10 (66.7%)	5 (11.6%)	**0.001**
L31M	6 (10.4%)	6 (40.0%)	0	**0.001**
H58C	1 (1.7%)	1 (6.7%)	0	0.259
H58P	1 (1.7%)	1 (6.7%)	0	0.259
Y93C	1 (1.7%)	0	1 (2.3%)	0.437
Q30R	1 (1.7%)	1 (6.7%)	0	0.259
M28V	1 (1.7%)	0	1 (2.3%)	0.437
Q30H, Y93H	4 (7.0%)	1 (6.7%)	3 (7.0%)	0.967

SVR: Sustained virological response; RAS: Resistance-Associated Substitutions

However, when analyzing NS5A RAS frequency by clade, a higher prevalence of RAS was found in clade I than in clade II (66.7% vs. 11.6%) with an associated statistical significance (P=0.001). The most frequent RAS in clade I was L31M (40%), which was

completely absent in clade II (p = 0.001). The clustering of the mutation suggests its forward transmission. The combination Q30H + Y93H (7%) was the most prevalent in clade II (7%) but was also present in clade I (6,7%). For this and all the remaining RAS, except for L31M as mentioned above, the statistical association to a specific clade was not found to be significant (p > 0.05) (table 3.8). Despite an association of L31M NS5A RAS to clade I, treatment outcome is not affected by HCV clade type (table 3.8).

Even when analyzing relapses and SVR cases stratified by subclades (table 3.9), and despite SVR occurring at a higher percentage in clade II (75.5%) when compared to the subclades I Y and I X, the outcome still did not show any dependence on the subclade 1Y and 1X or clade II strains (p > 0.05).

Table 3.9 SVR and relapsed cases and their stratification by clades/subclades.

Outcome	Total of samples	Subclade I Y	Subclade I X	Clade II
Relapse (n, %)	5 (8.6%)	0	2 (40.0%)	3 (60.0%)
SVR (n, %)	53 (91.4%)	9 (17.0%)	5 (9.4%)	40 (75.5%)
		P = 0.416	**P = 0.078**	**P = 0.596**

3.4.2 The specific presence of L31M in subclade 1Y revealed a statistical significance

The strains with L31M mutation cluster together in subclade Y, suggesting forward transmission of this mutation. Most clade I strains with RAS belonged to subclade Y (n=8; 88.9%), while a minority (n=2; 33.3%) revealed to belong to subclade X (Figure 3.16). Likewise, L31M, H58C and H58P were found only in Y strains, while Q30H showed to be present only in subclade X strains (Figure 3.16) and Q30H+Y93H was mainly present in clade II but also in one sample of subclade I X. When comparing the two subclades of clade I, only L31M specific presence in subclade 1Y was significantly statistical supported (p = 0.000), while the presence of the remaining RAS did not seem to present a significant statistical difference between subclades [H58C (p = 0.148), H58P (p = 0.148), Y93C (p = 0.739), Q30R (p = 0.096), M128V (p = 0.739) and Q30H+Y93H (p = 0.379)].
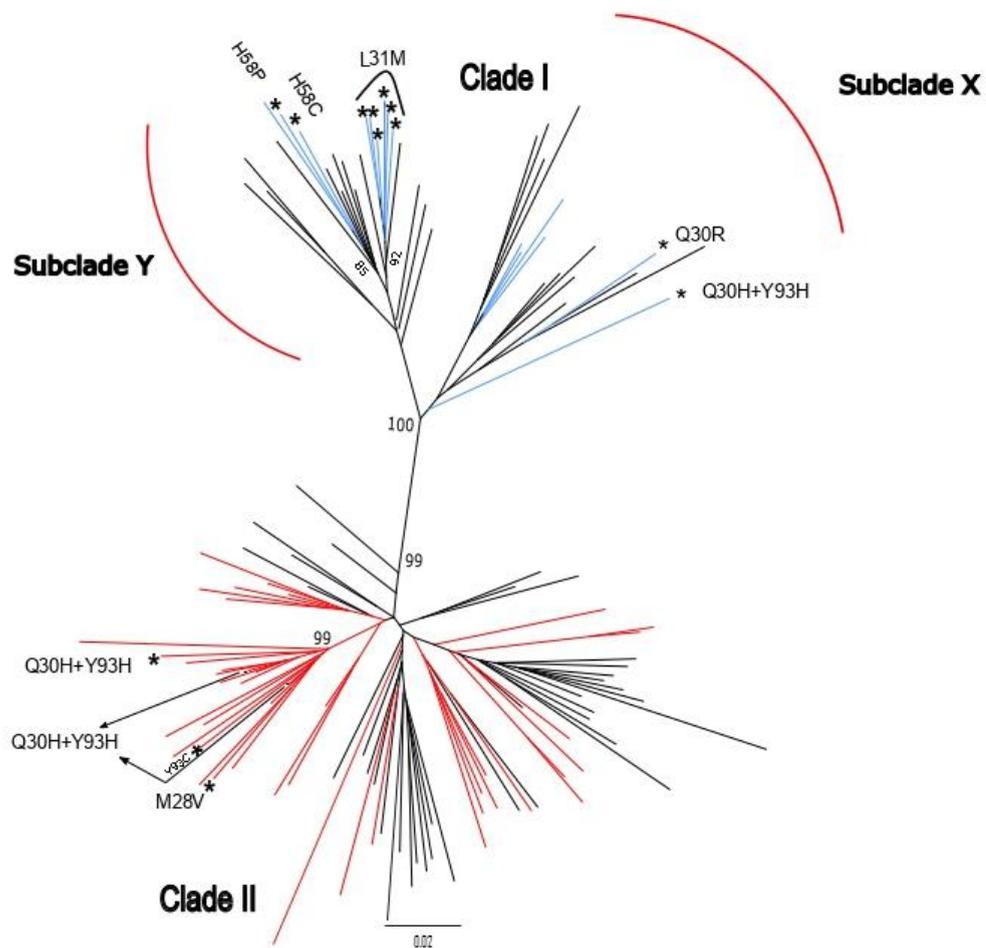
Figure 3. 16 | Phylogenetic tree showing HCV clade I (and respective subclades) and clade II strains.
Unrooted Maximum Likelihood phylogenetic tree showing Portuguese HCV strains with RAS marked with
*. Each specific RAS is pointed out in front of the respective strain. Clades I strains are visible as blue lines
and clade II strains as red lines. The red round lines show the subclades of clade I. The most relevant
bootstrap support values are marked next to the respective internal nodes.

3.4.3 Relapsing occurs in the absence of mutations

Among the five relapsed patients, one presented the NS5A RAS pattern Q30H+Y93H.
Table 3.10), while the remaining four relapsed patients had a complete absence of RAS.
The pattern Q30H+Y93H was present in 21.4% (n=3) of the SVR cases with RAS (Table
3.10). However, the outcome did not reveal any dependence on Q30H-Y93H RAS pattern
(p = 0.315) or on any NS5A RAS in general (p = 0.748).

Table 3.10 Percentage of samples showing NS5A RAS by clade within SVR and relapsed cases.

RAS	SVR % (N)	Clade % (N) I Y	I X	II	Relapses N	Clade % (N) I Y	I X	II
L31M	42.9% (6)	75.0% (6)						
Q30H, Y93H	21.4% (3)			60% (3)	100% (1)		100% (1)	
H58C	7.1% (1)	12.5% (1)						
H58P	7.1% (1)	12.5% (1)						
Q30R	7.1% (1)		100% (1)					
M28V	7.1% (1)			20% (1)				
Y93C	7.1% (1)			20% (1)				
Total	100% (14)	100% (8)	100% (1)	100% (5)	100% (1)		100% (1)	

In addition, most of the patients presenting NS5A RAS revealed a SVR after treatment [n=14; (26,4%)] and independently of the presence or absence of RAS, undetectable viral load was achieved at week 12 in these patients.

DISCUSSION AND CONCLUSIONS

The present work is the first report to investigate the epidemic and evolutionary history and phylogeography of Hepatitis B and C viruses in Portugal.

The specific findings of each individual study, described in the previous chapter, are discussed here in detail and in addition studies limitations, major conclusions which we consider arise from this research and future research perspectives will also be addressed.

4.1 Phylogeography of HBV in Portugal

4.1.1 Origin of HBV

Our work was the first to reconstruct the spatial and temporal dynamics of HBV in Portugal using a Bayesian framework.

While previous studies recurrently reported HBV genotype D as the most prevalent in Portugal, most of them were done until 2008 or 2009. Furthermore, those done more recently, until 2013 or 2014, were performed in the north or center of the country or with samples collected from specific vulnerable groups. such as prisoners, pregnant women, or blood donors (39, 106).

On the other hand, methodologically, none of these earlier papers used phylogenetic analysis to classify the HBV strains. Most genotyping tests were performed with commercial methods such as INNO-LiPA HBV (Fujirebio; Tokyo, Japan) or Trugene Genotyping kits (Siemens Healthineers; Erlangen, Alemanha), which are more practical to perform in a routine analysis laboratory. INNOLiPA requires the hybridization of PCR products from *pol* gene domains B to C with HBV genotype-specific probes attached to nitrocellulose strips. The results of HBV genotyping are interpreted using a kit interpretation chart and can be quite difficult, depending on the number and density of the bands shown on each sample strip. However, HBV mutations may complicate hybridization of PCR products to the probes leading to untypeable band patterns (198). The Trugene Kit is a direct sequencing-based method whose software acquires forward and reverse sequences data in real time, which are then assembled to the final sequence and aligned with reference sequences representative of each genotype, that correspond to the surface antigen and polymerase regions. This basic genotype classification does not specify the subgenotype and relies on a simple comparison of sequence homology based on the percentage of identical bases within the sequenced region between the sample and the reference sequences stored in the software (199). The advantage of our approach, the

phylogenetic analysis, is to incorporate the evolutionary history into the genotyping of strains. This is in fact the most accepted approach for genotyping viruses and the most widely used. It is more accurate, relies on the establishment of evolutionary relationships between viral strains and based on phylogenetic analysis, allows to understand how they diverged from a common ancestor.

The current phylogenetic analysis, performed with HBV strains isolated from the general population followed in the largest hospital of Lisbon between 2005 until 2012, shows that genotype A strains represent a higher percentage of strains compared to genotype D, and that genotype E is the least common genotype. Our results might reflect recent changes in population migratory patterns in recent years. In particular, the entry of Portugal in the European Union may have contributed to increase of immigrants and/or to the entry of migrants from different regions, which became more intense since the first decade of 21st century (200), specially from European countries (201), followed by African and American countries (202, 203). American emigration to Portugal overlapped African countries since 2009. Therefore, during the period of samples collection in this study, there was a high foreign influx of migrants into Portugal specifically coming from Cape Verde and Brazil but also from Ukraine. (204-207). Coincidently, HBV genotypes A (A1 and A2) and E prevail in Cape Verde (208) and HBV/A1 is also the predominant subgenotype in Brazil, while HBV/A2 is the second most predominant subgenotype (151). This may explain the higher prevalence of genotype A than genotypes D in Portugal, when compared to earlier reports (106).

An estimated mean substitution rate of 9.45×10^{-5} s/s/y (95% HPD: 4.22×10^{-5} - 1.43×10^{-4}) was obtained, which agrees with other recent HBV genome studies that also describe evolutionary rates between 10^{-5} - 10^{-4} s/s/y (91, 151, 209, 210). Molecular-clock-based analysis has been used to reconstruct the evolution of HBV as well as dating its origin, but the results have proven to be still unclear and controversial (135-137). Some studies have dated the origin of HBV to the 16th century (141) or to the 17th century (211) while other to 7000 (143) or even to 33600 years ago (139), resulting in an enormous uncertainty about its timing of origin. Our data traced back the origin of HBV to the present region of France around year 886 (medieval age) with a HPD interval between the years 414 B.C

(Ancient Age) and 1514 (Modern Age). In fact, there is recent evidence of the presence of HBV in Medieval and Neolithic human remains from Europe, further supporting that HBV is an ancient virus in this continent (143) and reinforcing the possibility that the origin of HBV might be in Europe.

Discrepancies and the large confidence intervals around our estimated dates of origin of HBV can be explained by the larger length of older branches. If HBV is truly an ancient virus as all the above evidence recently indicated (135, 137, 139, 143, 211), it explains the larger length of older branches in our HBV MCC tree. In addition, the fact that we could not include such old reference samples increased the uncertainty around the estimated average year for the MRCA of each internal node, especially near the root. The sample size is inversely proportional to the degree of uncertainty, so the smaller the ancient reference sample size, the larger the 95% HPD width, which indicates a higher degree of uncertainty (212).

4.1.2 Phylogeography of HBV/A1

In agreement with Kramvis and Paraskevis (147), our study traced back the origin of HBV/A1 to Africa. However, contrarily to Kramvis' suggestion that pointed the origin of subgenotype A1 to Southern Africa, or even another theory that purposed East Africa as the origin (146), our data raises the hypothesis that the origin was in the North Africa, around year 1711 [95% HPD: 1416 - 1855]. HBV/A1 evolution seems to have occurred not just along the North of the continent and Chad, from where it appears to have been exported directly to Brazil and India (clade I), but also in Kenya. From this region, it dispersed across Eastern, Central and Southern Africa (clade II), and then to Europe via Portugal and Belgium (Annex 4). These two distinct clades correspond to the previously described Asian-American and African clades, respectively (147).

Most HBV strains from Portugal cluster in Asian-American clade (clade I), which has also been described in Brazil and Cape Verde (151, 208). Cape Verde is an African archipelago and Brazil received in ancient times many African slaves, and for these reasons it has not been easy to understand why the HBV strains from both countries are not grouped in the African clade. A previous study hypothesized that an alternative route to the slave trade could explain the fact that Brazilian HBV sequences clustered in the Asian-American clade instead of in the African clade, with a possible importation from

East Africa or Asia by merchants in the middle of the XIX century (151). However, our results suggest instead that HBV/A1 dispersed directly from North Africa to both regions, Brazil and India, in a period when Portugal was strongly present in both territories and participating in the slave trade especially to Brazil. Our temporal tree revealed that Brazil seemingly exported its strains to other American countries, namely Uruguay, Panama, Haiti, Cuba, USA, and Argentina, while India exported its strains to Japan, China, Bangladesh, and Philippines (see this detail in annex 4). Both countries contributed to the spread of the strains and probably to the emergence of the Asian-American clade. Portugal seems to have imported most of the HBV/A1 strains from Brazil near the beginning of XX century, and then exported it to France, Italy, Netherlands, Germany and to Cape Verde, at this time still a Portuguese colony. In fact, the Portuguese role in exporting HBV/A1 into Cape Verde may explain why Cape Verdean HBV/A1 strains are more associated with the Asian-American clade, in particular with Brazilian strains, than with the African clade as would be expected (208).

4.1.3 Phylogeography of HBV/A2

It has recently been suggested that the ancestral HBV strain of the A2 genotype emerged in Europe during the bronze age (3300 to 1200 BC) and that it originated from other genotype A strains from central Asia through the interaction of the Asian steppe populations with the European populations (145). The time span of our reference sequences is not as large as in the study that suggested this and that can explain the fact that our data suggests a more recent origin of HBV/A2. The molecular clock models used are sensitive to bias caused by the fact that reference sequences do not include a larger sampling period. As previously explained, the smaller the ancient reference sample size, the larger the 95% HPD width, which would cause a higher degree of uncertainty (212). Yet, our phylogeography analyses are coherent with their results, also suggesting Europe as the place of origin. Accordingly, to our results, HBV/A2 emerged between the middle of XVI and the end of XIX centuries in the region of present Belgium, which seems to have played a main role in the spread of HBV/A2 in Europe, exporting this subgenotype to Portugal more than once. A relevant role was also played by Portugal in spreading HBV/A2 outside of the European continent, mainly to Cape Verde between the beginning of the XIX and the middle of XX centuries. A2 strains from Cape Verde were previously

described to group into a single cluster quite separate from the HBV/A2 strains of the other geographic regions (continental Africa, Asia, the Americas or Europe), with the exception of a strain from Poland (208), a European country to which Portugal also seemingly exported HBV/A2 via France. Our results provide the first indication that Portugal was the source of HBV/A2 isolates found in Cape Verde.

4.1.4 Phylogeography of HBV/D3 and HBV/D4

On the other hand, our Bayesian analysis suggests that HBV/D may have emerged earlier than HBV/A, around the year of 1501. Also here, the results indicate a confidence interval that covers a vast period from 922 to 1790. Our study points to a putative origin of HBV/D in France, while previous phylogeographic analysis indicated North Africa/Middle East (213), Southern Europe, Central/Eastern Europe, Syria or Martinique (91) as the putative origin of HBV/D. The lack of HBV/D sequences from more diversified locations in the world and from consistent genomic regions is probably limiting consensus conclusions in phylogeographic studies.

In a phylogeographic study, genetic data is integrated with geographical and temporal sampling information. Therefore, it is essential to gather reference sequences with specific traits, mainly the genotypes or subgenotypes of interest, date of collection and place of origin. The collection of these samples according to these criteria would guarantee dense sampling. For this reason, it is necessary to be extra careful when choosing these sequences, since the non-random nature of these sequences can lead to the absence of the complete genetic diversity in the set of sequences. The major limitation we came across in this study, regarding the reference sequences of subgenotypes D3 and D4 retrieved from GenBank, is that the vast majority of those that had information about date of collection and place of origin came from studies carried out in Brazil and the North of America. As such, continuous space phylogeographic inference can be negatively affected by sampling biases, such as sampling being focused in certain areas over others, leading to potentially mis-inferred ancestral node locations, up to completely excluding the true origin of outbreaks with complete confidence (214). As an attempt to reduce the effects of sampling Bias, we made an additional effort to exhaustively search for sequences from smaller studies performed in different parts of the world and in different time periods so that the range and distribution of collected samples would reflect the

geographical distribution of HBV subgenotypes D3 and D4, as well as to improve the temporal sampling information.

Portugal seems to have played a fundamental role in the dissemination of HBV/D3 and HBV/D4 during the period that covers a great wave of Portuguese emigration to the American continent in the late XIX and early XX centuries. During this period, there were great human migrations in different parts of the globe, but none can compare with the 44 to 52 million Europeans that crossed the Atlantic to the American continent between 1815 and 1914 (215). This was caused by the European economic crisis and the labor crisis caused by the progressive cessation of slave traffic worldwide during the second half of XIX century. Many Portuguese from the continental region emigrated to Brazil, not only because of the language but also for religious affinity. On the other hand, Portuguese from Azores emigrated mostly to the United States (215).

All Portuguese HBV/D4 sequences clustered together suggesting a single introduction of this subgenotype in our country via France. This Portuguese cluster seems to be the most likely and almost exclusive origin of HBV/D4 infections in Brazil, whose sequences also cluster monophyletically with just one exception. This monophyletic cluster of D4 Brazilian sequences was previously described, and a single introduction of D4 in Brazil had also been proposed by the authors with a probable time of introduction of 1848 (CI: 1062-1946), consistently with our study (91). However, spatial origin was pointed at Martinique, which could be due to sampling bias and the authors considered incompatible with epidemiological and historical data of HBV in the Americas (91). Our results help to further elucidate on these previous results, now showing that Portugal is at the origin of Brazilian D4 strains with its most recent common ancestor (tMRCA) dated around 1894 (CI:1788 to 1944) which coincides with the period of massive emigration from Portugal to Brazil by the end of XIX century (215). As the confidence interval still covers the end of the XVIII century, it is not possible to exclude that the slave trade factor could have contributed to the introduction of this subgenotype in Brazil. However, the parallel exportation of HBV/D4 from Portugal to France (1928) and Germany (1955), which were secondary emigration destination of Portuguese people in the first half of last century, further support the median estimated date and not the extremes of the confidence interval, suggesting the first theory of Portuguese emigration at the end of the XIX century.

Subgenotype D3, introduced in Portugal perhaps a few years later than the HBV/D4 (late of XIX century) but also with a probable origin in France, seems to have been exported with Portuguese population together with the Italians and French to Brazil in the beginning of the XX century, around 1909 (1814 to 1952). HBV/D3 was previously described to have been introduced in Brazil around 1799 (1615 to 1976) by Southern European people (91), especially during the mass European emigration from different countries encouraged by the Brazilian government campaigns, such as Italians, Portuguese, Spanish, Japanese, Germans, Lebanese and Syrians (152). The Italians, who were divided between emigration to Europe and Argentina, only came to Brazil in large groups in the last two decades of the XIX century (215), probably introducing genotype D3 in the country as previously proposed (151). Interestingly, this agrees with our results that indicate that the tMRCA is near 1895 for a group of HBV/D3 Brazilian sequences with origin in an Italian cluster. From the moment the Portuguese people first arrived in South American lands there was always a flow of Portuguese people to the region that became Brazil. Nonetheless, the largest flow of Portuguese emigration started in late XIX and the beginning of XX centuries (215), which agrees with our estimated tMRCA of 1909 for the introduction of Portuguese HBV/D3 in Brazil. The confidence interval of this tMRCA has a lower limit of 1814, which is already very close to Brazil's independence (in 1822) and consequently to the end of the slave trade by the Portuguese people. This gives more support to the possibility of HBV/D3 introduction in Brazil by the Portuguese to have occurred during the emigration flow in XIX - XX centuries rather than with the slave trade. Although with less intensity, Portugal also seems to have introduced HBV/D3 subgenotype in the USA, Spain and France, which agrees with other secondary migratory destinations of Portuguese people (215).

4.1.5 Phylogeography of HBV/E

Genotype E has been assumed to have a more recent origin probably in the present region of Nigeria around the last 130 to 200 years ago, which is supported by its low genetic diversity between sequences of this genotype, characteristic of a shorter evolutionary history (94, 154). The tMRCA found in our study, 1818, is in full agreement with the previous fact. However, contrarily to the theories of those authors, our data supports France as the place of origin of HBV/E, instead of Africa. One possibility for this disparity

is the fact that some HBV/E sequences available in the databases do not indicate sample collection date and geographical origin and therefore could not be used in our study, decreasing the availability of the sequences from different regions described as being in the origin of this genotype. This indicated a potential sampling bias that can cause an erroneous estimation of the place of origin of HBV/E.

Despite the low genetic variability, HBV/E is hyperendemic throughout West Africa, reason why a rapid spread in countries along the West African coast has been proposed, from Guinea to the Central African Republic and subsequently towards Eastern and Southern countries, such as Sudan, Angola and the Democratic Republic of the Congo (94, 154). As revised by Muller *et al* in 2009, an explanation for this phenomenon would be iatrogenic transmission in mass-injection campaigns with unsafe injections by the colonial governments (155), which has also been proposed as contributing to the dissemination of other infectious diseases such as HIV (216). Interestingly, Portugal seems to have imported HBV/E from Spain in the beginning of XX century. Our estimated dates of introduction of HBV/E in Portugal fall precisely within the period between the 1880s and 1930s, previously described as the most significant expansion period of this genotype (94). While the Portuguese role was markedly important for the spread of the HBV/A1 and HBV/A2 during the slave trade, a clear difference is evident in this case since Portuguese sequences are not causing outstanding exportations to other countries. This agrees with the absence of HBV/E in descendants of African slaves in South America that was assumed to indicate that this genotype was most likely not present in the West African population in the period of slave trade (94, 148, 149). The absence of the HBV/E from the South American continent also indicates that Europeans who emigrated to the Americas did not introduce it there. This fact reinforces the idea that during the period of emigration from Europe to the American continent, in the end of the XIX century and the beginning of the XX century, HBV/E was probably still absent or weakly incident in Europe. This explanation is also in line with the reduced Portuguese exportation of HBV/E to other countries. Furthermore, these findings can also indicate an even more recent origin of the genotype than the one estimated in ours and in previous studies.

In summary, Portugal seems to have actively contributed to the dissemination of HBV genotypes around the end of XIX and beginning of XX century, except for HBV/E for

which contribution to its dissemination was more modest. Considering historical facts, we believe that the export of HBV by the Portuguese was due to migratory factors that reached the whole of Europe in the period in question. However, in this study there was one fact that stood out in relation to HBV/A1. After its emergence on the African continent, this subgenotype dispersed simultaneously to Brazil and India around 1869 with a time interval that spans the period from 1732 to 1923. Historical events show the presence of the Portuguese people in the three continents during the period in question, a period that was also marked by its intense participation in the African slave trade. All these facts together raise the strong hypothesis that Portuguese contributed to the spread of this genotype during the time when they participated in the African slave trade.

Some problems and limitations were faced along this study. The main limitation was the inevitable exclusion from the analysis of sequences that were present in GenBank due to not gathering all the necessary conditions to be used in the spatiotemporal reconstruction as referred above for HBV/D strains. Many of the public sequences corresponding to the region of the *pol* gene under analysis did not have a date of sampling or country of origin and had to be excluded. For sequences that fulfilled the full requirements, many belonged to the same study or were collected in the same country. It was challenging, difficult and very hardworking to construct an unbiased dataset with reference sequences reflecting the genetic diversity and geographical distribution of HBV and at the same time presenting temporal information. We consider that it is urgent to take some measures to facilitate phylogeographic studies. On the one hand, it would be important to standardize the sequenced genomic regions of HBV and to require that authors sharing sequences provide the date of collection and country of origin, so that there is a greater number of sequences that can be analyzed in this type of study. On the other hand, it would be important to make HBV sequences from countries where the virus is present but less studied available. We found that HBV subgenotypes from some regions of the globe such as Angola, Guinea Bissau, Mozambique, with which Portugal maintained a close relationship, were underrepresented and therefore it was extremely difficult to equally represent them in our dataset.

In addition, GenBank also doesn't have sequences thousands of years old, so the representation of ancient times is absent. For this reason, the 95% HPD intervals for the

oldest branches are width, which translates into a large uncertainty of the estimated year of origin of each subgenotype.

Another problem that was complicated to solve was the difficult convergence of some parameters in the analyses performed in BEAST software. We believe this is likely attributable to the complex biology of HBV genome, which has a small size and has extensive overlapping reading frames resulting in high variability of substitution rates across the genome, for example, between the non-overlapping and overlapping coding regions (217, 218). The pol genomic region analyzed in this study overlaps with s gene in about 80% of its length while the remaining fragment is a non-overlapping region. It is possible that the difficult convergence of BEAST results is caused by the small non-overlapping part which was described as having a higher estimated mutation rate (218). In addition, we decided to analyze all the different HBV genotypes in the same dataset, because the convergence of results was worse when datasets were separated in different datasets. While it is described that each genotype evolves at different rates, the usage of a relaxed clock model should allow to incorporate for such rate variations (217, 218).

In fact, we were able to obtain results only when using an uncorrelated relaxed molecular clock, which contrarily to the strict clock model assumes that every branch in a phylogenetic tree evolves according to the same evolutionary rate, letting the evolutionary rate across branches change abruptly. Besides testing HBV demographic history with expansion and exponential growth models as priors, we took a more flexible approach using non-parametric coalescent priors such as Skygrid and Skyline to capture a possible more complex demographic trajectory as can be expected from a virus as old as HBV. And in fact, only the Bayesian Skyline converged, maybe because it allows for the presence of several periods of constant demographic growth, with instantaneous changes of effective population size and demographic growth rate between consecutive periods (219).

4.2 Epidemic history of HCV genotypes and subtypes in Portugal

Phylogenetic and phylodynamic analysis of NS5B sequences obtained from HCV infected patients were applied to make the first characterization of the origin, diversity, and epidemiologic history of HCV in Portugal.

This study was characterized by a large number of samples gathered from patients that had received a recent diagnosis of HCV infection and whose disease was starting to be monitored. We consider this sample is more appropriate to reflect a good picture of the HCV subtypes that circulate in the Portuguese population than, for example, patients gathered with the aim of changing therapy. In such case, sampling could be biased to represent subtypes more frequently failing therapy.

Although the 230 samples collected belong to patients followed in the Hepatology department of a single hospital in Portugal, Santa Maria Hospital in Lisbon, we think that it likely reflects the national reality, since this is one of the largest Hepatology departments that assist HCV infected patients in the country.

Some clinical information was not available for all patients and that did not facilitate some aspects we would like to have studied. For example, trying to establish an association between the subtypes and their probable route of transmission was a difficult task.

Furthermore, the epidemiological history was not reconstructed unless for the most frequent subtypes, because the evolutionary estimates of the rarest subtypes could be affected by their small representativeness.

4.2.1 HCV genotypes and subtypes circulating in Portugal and their epidemiologic history

Similarly to what has been reported in previous studies, the most prevalent strains in Portugal were HCV/1a, representing almost half of the sequences analyzed, followed by HCV/3a and HCV/1b (100-103). These genotypes are responsible for most of the HCV cases globally (75, 220, 221). According to our estimates, HCV/1b has caused a slowly growing but extended epidemic from the 1930s to the 1990s. This epidemic preceded all other genotypes studied in this work, by three decades and its features are consistent with widespread unsafe needle practices at the beginning of the 20th century and contaminated blood transfusions which stopped only after the discovery of HCV in 1989 and the subsequent screening of all blood donations for HCV (222, 223). The increase of

intravenous drug use in the 1960s has likely sparked the second major HCV epidemic in Portugal which was caused by HCV/3a and went through the beginning of the 1980s. The fastest growing HCV epidemic in Portugal occurred during the 1980s and 1990s and was caused by HCV/1a. It was likely driven by PWID which increased dramatically in the late 1980s and 1990s (up to 1% of the Portuguese population) due to the rapid social and cultural changes that followed the revolution of 1974 and the increasing availability of heroin (224-226).

A relatively high (16.1%) prevalence of HCV/4 was observed among the study population, that is about three-fold higher than the estimated prevalence for Western Europe (220, 221). Similar results were obtained recently in HCV-infected inmates in Portugal (13% HCV/4) (104). In addition to HCV/4a, the most prevalent in this study group, and HCV/4b and HCV/4d that have already been reported in Portugal (103, 104, 227), one isolate classified as HCV/4f and one as HCV/4k were also identified among the samples. To the extent of our knowledge, this is the first report of these variants in Portugal. Globally, HCV/4 has been more prevalent in some countries of Northern and Central Africa and the Middle East (174), but there are indications that HCV/4 is becoming increasingly prevalent in Europe, mainly among PWID (85, 103, 228). It was interesting to find that unlike HCV/1 and HCV/3 that caused endemic localized infections during decades before becoming generalized, HCV/4a seems to have become epidemic almost immediately after being introduced in Portugal, likely in 1980, and this epidemic may still be growing at a relatively fast rate. The reason for this is unclear as the mode of transmission was known only for 27% of our patients. HCV/4a dominates the HCV epidemic in Egypt and the main risk factor for its origin and transmission has been unsafe medical practices (134). Like HCV/4a, the epidemic history of HCV/4d indicates a recent introduction and a progressive epidemic in the Portuguese population. In Portugal, healthcare related HCV infections are exceedingly rare and HCV/4a and HCV/4d have been found in relatively high frequencies (11.3% and 12.1%, respectively) in PWID analyzed in a similar period (2008–2009) in Lisbon (103). In addition, 83.3% of HCV/4a (n = 6) and 75% of HCV/4d (n = 4) infected patients for which we have information on transmission risk were PWIDs. We, therefore, believe that injecting drug use is the main mode of transmission of HCV/4 in Portugal.

The prevalence of HCV/2 in the study group was only 3.2%, about one-third lower than previously reported data from Western Europe (221). In particular, the prevalence of HCV/2a, generally considered as an epidemic subtype, was less than 1%. A similar low prevalence was also observed for HCV/2c, in contrast to the high frequency (37.5%) previously reported among a limited number (n = 64) of liver biopsies at the Portuguese tertiary Hospital Santa Maria, the same healthcare setting attended by this study group (101). Concerning the other less prevalent HCV variants identified in this study, one individual harbored HCV/1g. To the best of present knowledge, this is the first report of this subtype in Portugal. In Europe, the first partial sequences of HCV/1g were derived in 1994–1995 from HCV chronically-infected German patients who were probably immigrants from Egypt and Sudan (229). More recently, the first complete genome of a GT1g isolate derived from a Spanish patient was published (230).

4.2.2 Prevalence of HCV/1a clades in Portugal

Recent studies have highlighted distinct geographic distributions of the two clades of HCV/1a, with clade I being more prevalent in the United States and both clades equally distributed in Europe (73, 231). Furthermore, a recent in-depth phylogenetic analysis has found three distinct sub-clades within clade I (232). In the present study, that represents the first assessment of HCV/1a clades in Portugal, we found a four-fold higher prevalence of clade II than clade I. Interestingly, the prevalence of clade I among the patients included in the present study (21%) was similar to that recently observed in DAA-naïve subjects in Spain (233) but it was significantly lower than in other European countries (e.g. 48% in Italy and 67% in France) (231). The prevalence of clade II was also like the one found in Spain. This may indicate a difference in the temporal spread of the HCV/1a clades in both Iberian countries as compared to other European countries where the two clades are equally distributed. This Portugal/Spain similarity in clades' frequencies, and difference in relation to what is described for other European countries, may indicate a relationship in the evolutionary history of clades in the Iberian Peninsula.

In summary, the most frequent HCV subtypes are HCV/1a, HCV/3a, HCV/1b, HCV/4a and HCV/4d, respectively. The first epidemic to expand in Portugal was caused by HCV/1b around 1930 likely due to unsafe needle practices and contaminated blood

transfusions, followed by HCV/3a around the 1960s due to the increase of intravenous drug use and HCV/1a during the 1980s driven by the PWID epidemic. Finally, the most recent subtypes to entry in Portugal were HCV/4a and HCV/4d, and while the HCV/4d epidemics seems to have reached a plateau, HCV/4a epidemics seems to be still growing. Among the most frequent HCV subtype, HCV/1a, there are two clades circulating in Portugal, that showed a different prevalence between them and from other countries of Europe, excluding Spain whose clades prevalence are very similar to Portugal. The different prevalence between clade I and II in Portugal makes us question whether this could indicate that they had a different temporal spread. Furthermore, the fact that the prevalence is similar between Portugal and Spain but different in relation to other European countries, makes us question whether this is an indication of a relationship in the evolutionary history of these clades in the Iberian Peninsula.

4.3 Evolutionary history and phylogeography of Clades I and II of HCV subtype 1a in Portugal

4.3.1 HCV/1a clades prevalence and phylogenetic analysis

Even using a different group of infected individuals and analyzing a different gene, a 2.9-fold higher prevalence of HCV/1a clade II strains than clade I (74,14% vs 25.86%) was observed, which agrees with the previous work done (197). Unlike the former study that was performed with NS5B sequences, in this study lineage assignment by phylogenetic analysis was consistent with the Geno2pheno[HCV] algorithm results, indicating that Geno2pheno[HCV] might perform better with *NS5A* sequences or might be more efficient in the classification of longer DNA fragments due the presence of more genetic information. Interestingly, a statistically significant higher prevalence of clade II in men and clade I in women was found which, in the absence of potential sampling bias, can suggest different transmission pathways for these clades. To our knowledge this association has never been described or studied.

4.3.2 Origin of clades I and II

Our molecular clock analyses indicated similar dates of origin (1934) for both temporal trees (clade I and clade II), which may represent the timing of divergence of clades I and II, with a confidence interval that vary between 1918 and 1947. De Luca *et al* (231) estimated a later divergence date of both clades [1964 (95% HPD 1941 – 1976)], which reflects a difference of 3 decades when compared to the estimate of our study. However, our sequence datasets have a more diverse group of reference sequences that are more representative of different regions and cover a wide temporal span. As such, we suggest that this difference can be explain by a lack of representativity of the full genetic diversity in each clade in De Luca *et al* study, that results in the estimation of an erroneous tMRCA of each clade. On the other hand, another recent study in Spain, estimated early introduction dates for both clades in that country (Clade II: 1912; Clade I:1957), one of them earlier than the tMRCA estimated herein, which reinforces that clades I and II of subtype 1a may in fact have diverged earlier than the previously suggested date of 1964 (234).

In this work, we also found that the location where each clade emerged differs. United

States of America and France were indicated as the ancestral state location for the origin of clade I and of clade II, respectively. According to what has been described in the literature, despite the fact that the origin of HCV genotype 1 appears to have been in West Africa or Southeast Asia due to genetic patterns that suggest a long period of endemicity (235, 236), subtype 1a in particular seems to have diverged from 1b in USA, continuing its evolution in other high income countries and only spreading later to low and middle income countries (177). For this reason, it is possible that the divergence of the two clades of subtype 1a also occurred in high income countries, as our results suggest. In addition, it is described that about 73% of the subtype 1a sequences in the USA fall in clade I and that in Europe 67% of the sequences belong to Clade II (232), which further suggests that clades I and II have been present in the United States of America and France, respectively, for a long time.

4.3.3 Clades I and II emergence in Portugal and their dispersal routes

According to this study, the ancestor of HCV/1a strains circulating in Portugal dated back to 1955 (95% HPD: 1943-1965) which is not very far from the estimates of 1950 in the previous work about the Epidemic history of HCV genotypes and subtypes in Portugal (197). We now know that the first HCV/1a strain that was introduced in Portugal probably belonged to clade II and was imported from our Iberian neighbor country. In fact, Spain was the main introductory source of subtype 1a in Portugal, being the origin of 67% of our clade II strains and of 40% of those belonging to clade I. In addition, the introduction period of the two clades in Portugal via Spain, coincides with the period of exponential growth of each clade inside Spain (234), reinforcing our estimates. The multiple introductions of clade II via Spain, between 1955 and 1963, are not easy to explain. At that time Portugal lived under a dictatorial regime for about 30 years, which would only end in 1974, and which kept it isolated and away from external relations even within the Iberian Peninsula (237). On the other hand, Spain, which was also living in the core of a dictatorship, had closed itself to government relations with Portugal, and even during the wave of emigration that affected it due to the life difficulties that the population faced in the country, Portugal was never an elected destination for the Spanish people (238). Two effective transmission modes for the first cases could be drug use or blood transfusion. Concerning drug use, the whole movement of revolution associated with drugs use that

occurred in the 60s went unnoticed by the population at least until the end of the 70s, once Portugal was a very closed and conservative country due to the dictatorship in which it was immersed (239). Furthermore, drug use was more frequent in the context of soldiers in the African colonies, which could point to a different origin other than Spain for infectious agents related to this mode of transmission. As for transfusions, the registration of the first official blood donor in Portugal dates back to 1944 and at that time blood donations were not tested for potential infectious agents as they are today. However, during a long period, blood donors were usually some family members of hospital patients and mostly people with weak economic resources, very low physical vigor, often poorly fed beggars (240). The likelihood that these people have contributed to import HCV strains from a foreign country, such as Spain or France, seems also reduced. However, between 1961-1974, it is natural that these sources of blood collection were insufficient to treat Portuguese soldiers wounded in combat in the war that African colonies fought with Portugal for their independence (241). Although there are no records of blood collections coming from abroad to help with blood scarcity for transfusions, this could be a potential hypothesis for the importation of these clades to Portugal.

Historical facts show that in the 50s, 60s and early 70s, contact of Portuguese population with Spain was constant near the borders, not just between neighboring villages from both countries but specially due to the intense smuggling routes of food products that were scarce in both countries and also coffee, tobacco, sometimes some medicines (242). These networks that kept the two countries in contact by that time may have played a key role in HCV transmission. The introduction of subtype 1a in Portugal could have been through sexual route, which is not as effective as the needles sharing due to use of intravenous drugs and may explain why, following its introduction in 1955, this genotype remained unnoticed until the 80s (197) when the use of drugs finally had a great impact in Portugal causing an explosive growth of HCV/1a inside the country.

Regarding clade II in particular, France's strong contribution and Italy's modest participation in its importation to Portugal in the 1960s probably was likely due to the Portuguese emigrants, most of them illegal, who began to seek improvements in living conditions abroad, more frequently in France (243), but also in Germany and Switzerland (244, 245). This last country also received many Italian emigrants (246), thus establishing contact between them and Portuguese emigrants.

Likewise, HCV strains of clade I were introduced in Portugal in the second half of the 1960s, possibly also due to the Portuguese returns that clandestinely emigrated to the USA and France (245, 247). Spain also contributed significantly to the introduction of clade I around 1975 in a post-dictatorship era and France in 1995, a time when the circulation of people and drugs was already easier and usual.

The evolution dynamics of both clades in Spain was recently studied and revealed that clade II was the first to emerge in Spain in the beginning of twentieth century around 1912 followed by clade I several decades later near 1952 (234). Clade II went through exponential epidemic growth only in the 1950s and clade I in the 1970s (234) and this may explain why Spain firstly introduced clade II in Portugal in 1955 and clade I only two decades later, in 1975.

Portugal spread clade I earlier into Spain around 1967, 1969 and 1973 eventually contributing to the beginning of the exponential growth of this clade in Spain. Interestingly, Portugal transmitted clade I to Spain but never transmitted clade II, which may reflect different preferential modes of transmission for each clade. As discussed above in this section, the statistical analysis carried out in this study indicated an association between gender and type of clade. In fact, it was possible to observe that men seem to be infected with more strains of clade II than clade I. If the transmission of clade II may have in part something to do with a specific male behavior, should be further investigated.

4.3.4 Global distribution of HCV/1a clade I and clade II

Our study's data also reveal that, globally, after the emergence of both clades in the 1930s, it was clade II that was immediately and effectively transmitted, increasing the number of infections, and establishing itself until the mid-1970s, which would only happen with clade I between 1950 and 1980. All together, these results are in line with what has been described for subtype 1a since it is usually analyzed globally without separation of clades and generally reveals intense growth at the time when intravenous drug use began to be more frequent, between 1960 and the end of the 80s (177). The decline of clade 1 between 2009 and 2011 may reflect the successful results of the HCV transmission prevention programs, such as needles exchange, along the decade before 2009. The fact that clade II does not show a decline reinforces the theory that each clade may have a different

preferential transmission route. It is possible that both clades are transmitted with the same efficiency using infected needles, but it is also possible that clade II has a second route of transmission that is very effective for these strains and that it is not as effective for clade I. Since men revealed that they were mostly infected by clade II, it is possible to raise the hypothesis that the homosexual transmission route is specifically as effective for clade II as the use of infected needles, but not as effective for clade I. This would be a possible explanation for the global decline of clade I between 2009 and 2011, while clade II remained stable. Based on the assumption that this decline was due, for instance, to lower levels of needle sharing as a result of the prevention programs implemented, possibly the transmission of clade II was maintained stable because this alternative route kept the clade transmission going on.

4.3.5 Clade I revealed the presence of two subclades

Our phylogenetic approach revealed for the first time in Portugal, the division of clade I strains into 2 subclades. This division of clade I has been previously described (232), although not in 2 but in 3 structurally distinct subclades (A, B and C). Since in that study the accession numbers of the sequences of each subclade were not provided, it was not possible for us to compare our sequences with those, to assign the same classification to our subclades. For this reason, in this study the subclades found in clade I were only classified as X and Y. That study in particular (232) presented differences in the occurrence of a natural polymorphism, Q80K, known to confer resistance to Simeprevir in the *NS3* gene. The proportion of the major resistance mutation Q80K was 64% in subclade IA, 0% in subclade IB and 3% in subclade IC (232). This may indicate that different clades and subclades may differ in the patterns of acquisition of specific mutations, and this issue should be further investigated, reason why we made a first approach of that study in the last part of this work.

In summary, clade II was the first to be introduced in Portugal around 1955 via Spain and clade I arrived around a decade later, near 1965, via USA. Despite the different origins, Spain has always been the biggest source of strains of both clades to Portugal. This fact may explain the similar prevalence of the clades between both countries. Nonetheless, France has also made significant contributions. The first crossing of Portuguese boarders

by both clades is difficult to explain in terms of transmission route since it occurred in periods when Portugal was immersed in a dictatorship regime and therefore very closed to contact with other countries of the world even with Spain. However, sexual intercourse and contaminated blood transfusions seem plausible hypothesis.

Globally, clade II transmission never presented a decline as occurred between 2009 and 2011 with clade I. In addition, in this sample we found a statistical association between gender and clades. Together, these facts may indicate that clade II has an alternative route of spread not so effective to clade I, that maintains its transmission stable.

New studies should be done, once one of the limitations of this study is the small sample size of strains of clade I due to its frequency being lower in the Portuguese population. With larger sample sizes of both clades, the statistical conclusions obtained in this study could be stronger. However, the inclusion of a large dataset of reference sequences was an approach to partly circumvent this limitation. As discussed in the two previous parts of this study, another limitation was the inevitable exclusion of the phylogeographic analysis of some reference sequences that were present in GenBank, but that did not gather all the necessary information for the spatiotemporal reconstruction. Many sequences corresponding to the region of the *NS5A* gene under analysis did not have a date of sample collection or country of origin and had to be excluded. For this reason, some regions of the globe, such as from African countries with which Portugal maintained a close relationship, were underrepresented in the datasets of the studied genotypes clades, limiting the analysis. As such, the analysis of potential importations or exportations to and from such African countries is limited.

4.4 Characterization of NS5A RAS and treatment response for the two clades of HCV/1a

Portuguese HCV strains are divided into 2 clades based on phylogenetic analysis. In addition, clade I was found to be divided into 2 subclades which were classified as 1X and 1Y. Disparities between clades and subclades regarding the natural NS5A RAS distribution were here investigated.

4.4.1 Clade type is associated to the higher prevalence of NS5A RAS and the presence of L31M in subclade 1Y revealed a statistical significance

It was notorious the most frequent appearance of RAS in clade I and the exclusive presence of L31M in subclade 1Y. On the other hand, the few cases that revealed RAS in clade II, mostly presented the Q30H + Y93H pattern. Statistical analysis confirmed NS5A RAS are dependent on the clade strain, seeming to be more associated with the clade I, but further studies with larger sample's population are needed once it would also raise the number of mutations found, enabling a robust statistical test use specially in the case of subclades. The study of natural NS5A mutations occurrence from the perspective of subtype 1a clades and subclades has never been carried out before. Such differential mutation occurrences can indicate that knowing if an HCV strain belongs to clade I or II, and, within clade I, if it is classified into subclade X or Y, can be informative for choosing which is the best antiviral to prescribe for each patient of genotype 1a. It is a fact that the current recommendations of the European Association for the Study of the Liver (EASL) suggest starting treatment regimens without having to perform a resistance test and even genotype determination is not considered absolutely essential (248), since there are pan genotype treatments for HCV infection that prove to be very effective for most genotypes of the virus. However, there are situations in which identifying the genotype and performing a resistance test can be useful. For example, there are atypical subtypes of genotype 1 that have revealed a suboptimal SVR (249). Some of these genotypes such as, 1l, 4r, 3b, 3g, 6u, 6v, among others, have natural polymorphisms that confer an inherent resistance to NS5A inhibitors, resulting in frequent and unacceptable virologic failures. The identification of the viral genotype before initiating first-line therapy can be useful in regions where access to drugs or their high price make it necessary for the therapy to be targeted more specifically towards a genotype to ensure its effectiveness with greater

certainty or also in cases of virological failure and treatment optimization (248). Besides this, in the situation of treatment failure and if there is a need to re-treat that patient who has previously been exposed to DAA, resistance testing is not absolutely necessary but can also help guide clinical decisions (248).

In this study, clade I/subclade Y and clade II strains presented most of the natural RAS, with L31M being the most frequent in the former and the pattern Q30H+Y93H the most common in the last. NS5A RASs are often detected at baseline in DAA-naïve patients (250). During host infection, HCV evolves rapidly, due to its high mutation rates and short replication time resulting in a genetically diverse viral population, i.e., quasispecies. However, upon transmission, there is a bottleneck that results in a marked reduction in the viral diversity, because only few of them will infect the new host (251). Each new host represents a novel environment for the virus where it has to prevail over natural barriers, such as host defenses, with fast adaptation process (251, 252). Random changes in the genetic profile of a population can happen due to chance (genetic drift) and in the process of adaptation any fitness advantageous genetic mutation may be positively selected or fixed, establishing a new population (252) that will become the most frequent and most likely to be forward transmitted to the next hosts. L31M has been shown to confer a higher replication capacity to HCV (253), as well as the Q30H+Y93H pattern (254). As such, we postulate that they may have become fixed after a genetic drift process in the sequence of HCV transmission. Another hypothesis is that this was a caused by a random drift effect, where by chance these strains were transmitted to other hosts and afterwards become fixed in the new host.

Using the 15% clinically relevant cut off associated to population sequencing in patients infected with subtype 1a, at least 1 RAS was found at baseline in 14% of the cases of European patients, being Q30H, L31M and Y93H the most frequent (250) which agrees with our results namely for L31M. High-level resistance is conferred in vitro by multiple substitutions at positions Q30, L31, and/or Y93 in the NS5A protein (255), being the specific substitutions that we found, Q30H, L31M and Y93H included. Treatment regimens with drug co-formulations that include Elbasvir or Ledipasvir may be affected by any of these mutations (256). According to an integrated analysis of phase 2 and 3 trials (257, 258) as well as in C-SALVAGE and C-EDGE TN trials (259, 260), these

mutations are often detected in subtype 1a viral strains of patients that revealed virologic failure. If the L31M and the pattern Q30H+Y93H are present in DAA naïve patients in either dominant or minority species, when treated with DAA, these resistant variants may become dominant. On the other hand, if these mutations were not present in naïve, their emergence could occur simply due to drug selective pressure, that can act differentially in viral strains that, due to genomic differences, may present also different 3D protein structures. These differences may lead to differential evolutionary pathways to emergence of drug resistance.

4.4.2 NS5A RAS presence does not affect treatment outcome

In our study, among the 58 patients, only 1 out of 5 relapsing cases revealed RAS, carrying the pattern Q30H+Y93H. This strain was classified as clade I/subclade X and was the unique of clade I that presented a pattern of RAS that is more frequent in clade II. It is not possible to conclude if this pattern is more critical in clade I than in clade II because there is not enough representativeness of these situations, and of relapsing situations in general, in this study, but further studies should explore this question. Other RAS described as present in cases of virologic failure are Y93C and M28V (250), which were also found in two different strains of clade II of this study, but both revealed SVR. However, they may affect, respectively Ledipasvir regimens and Dasabuvir, Ombitasvir and Elbasvir regimens (250). H58C and H58P that were present in only two strains of the clade I, subclade Y in this study and have also been reported to reduce susceptibility to NS5A polymerase inhibitors (248). Despite that, no relapse was detected in these patients, and it was not possible to find a significant statistical association between treatment outcome and the presence of RAS along this study.

In summary, clade I/subclade Y and clade II are undoubtedly the strains which presented most natural RAS, being the most frequent L31M and the pattern Q30H+Y93H, respectively. Thus, it is important to deepen the question whether belonging to clade I/Subclade 1Y or clade II is linked to the natural presence of certain mutations that reduce susceptibility or induce high resistance to NS5A viral polymerase inhibitors thus constituting a factor that has implications for treatments' failure in genotype 1a.

FINAL CONSIDERATIONS

This thesis has brought diverse new insights into the evolutionary dynamics and patterns of dissemination of hepatitis causing viruses. Specifically, we have studied the origin and dissemination of HBV worldwide and in Portugal and the phylogeography of HCV genotypes. Furthermore, we have explored the impact of such dissemination and evolutionary dynamics in the patterns of acquisition of drug resistance mutations.

Our approach to studying the introduction of HBV in Portugal not only revealed the origin of each viral genotype in our country, but also provided an interesting perspective on how Portugal may have played a key role in the spread of HBV around the world, especially to regions with which we have maintained a close relationship for a long time. Portugal seems to have imported most of the HBV genotypes from European countries except for HBV/A1 whose source seems to have been essentially Brazil and sporadically some African countries. Genotype A seems to have been increasing in Portugal, perhaps because of migratory movements after entry in the European Union and above all with the adhesion to the euro, which intensified the movement of people and goods over the last two decades. With our phylogeographic approach, the origin of the MRCA of HBV was pointed to Europe, more specifically the present-day region of France, about 1126 years ago, with a vast time span that varies between 498 and 2426 years. In fact, the origin of HBV remains immersed in doubt, as different studies point to different geographic origins in very different and sparse time periods. It is essential that the complete genome of more HBV strains, from as many countries as possible, be made available on GenBank and that they cover not only new geographic regions but also more remote time periods, as the possibility of HBV being an older virus than initially thought becomes increasingly concrete. Specifically, the isolation and sequencing of HBV strains from old paraffin specimens stored in different regions worldwide could be a future field of study. Specifically speaking of Portugal, since there are few published Portuguese HBV sequences, the regional and temporal diversity of the strains circulating at national level is very limited. In the future, it will be important to invest in new research studies to obtain sequences resulting from infections at various points in time and various regions of the country. This greater representation of HBV at the national level in space and in time would allow consolidating some evidence resulting from this work. Our results raise the exciting possibility that Portugal played a key role in the introduction of HBV/D4 and

HBV/A2 in Brazil and Cape Verde, respectively, clarifying what has remained unclear in the scientific literature until now. For this reason, it was very important to contribute to the publication of genomic sequences of HBV strains in GenBank. On the other hand, extend the study to the complete genome of the virus or at least to more than a single gene, with samples collected throughout the country and obtain more diverse HBV reference strains for the phylogeographic study would also overcome the geo-temporal limitations making it possible to substantiate these conclusions more solidly.

As for HCV, our study approach allowed us to reconstruct, for the first time, the history of the HCV epidemic in Portugal. Although HCV/1a is clearly the most prevalent genotype we now know that it was not the first to be introduced in Portugal. The epidemic began with HBV/1b in the 1930s probably due to unsafe needle handling practices in routine health care and transfusions with contaminated blood. It was only after HCV/3a emerged 3 decades later, with the start of intravenous drug use, that HBV/1a emerged in the 1980s associated with an increase in the number of PWID. As described in the literature, we also identified two clades in HCV/1a, but contrary to what is described for Europe, where both clades are distributed equally, clade II showed a prevalence four times higher than that of clade I in this study, which made us raise the question of a different temporal window and preferential patterns of transmission. Our results revealed that clade II entered Portugal via Spain, a decade before clade I was introduced via USA. However, Spain was the main contributor to both clades in our country. We found 67% of clade II and 40% of clade I strains in this study are related to HCV/1a strains from Spain. Our research approach brings very interesting new insights from genomic data and raises new research hypothesis. On the one hand, it showed a possible association between clade II and the male gender, which makes us raise the hypothesis that clades I and II have different transmission routes, maybe associated with different types of risk behaviors. On the other hand, it points to the hypothesis of preferential patterns of acquisition of RAS in different clades. Specifically, we raise the possibility that subclade Y of clade I may be associated with the presence of the L31M RAS in the NS5A gene and clade II with the Q30H+Y93H pattern. However, in our study only 5 cases of viral relapse were present. From our point of view, further research is essential to clarify whether the evidence described above is confirmed and independent of the limitations we faced in this work. A study with a larger number of samples belonging to each of the clades, and with associated

information about transmission route may help to confirm the association not only between clade and gender but also between the transmission route and clades. With a larger number of samples from each clade, the frequency of mutations found will possibly be higher, and if the number of relapse cases in the study is also increased, there will be more data that will allow a more reliable statistical association analysis about the type of clade, the presence of mutations and the outcome of the disease.

REFERENCES

1.   WHO. WHO Health Topics: Hepatitis. 2019. updated 1st September, 2019; Available from: https://www.who.int/health-topics/hepatitis#tab=tab_1.

2.   Jefferies M, Rauff B, Rashid H, Lam T, Rafiq S. Update on global epidemiology of viral hepatitis and preventive strategies. World journal of clinical cases. 2018;6(13):589-99. doi: 10.12998/wjcc.v6.i13.589

3.   Global Hepatitis Report. Published by: WHO; 2017. Available from: https://www.who.int/publications/i/item/global-hepatitis-report-2017

4.   Global progress report on HIV, viral hepatitis and sexually transmitted infections. Published   by: WHO;   2021.   Available   from: https://www.who.int/publications/i/item/9789240027077

5.   Global Health Sector Strategy on Viral Hepatitis; 2016–2021-Towards Ending Viral Hepatitis.   Published   by: WHO;   2016.   Available   from: https://apps.who.int/iris/bitstream/handle/10665/246177/WHO-HIV-2016.06-eng.pdf;jsessionid=73AFA6AD6B33CC58B8136A0D843E1B0F?sequence=1

6.   Doyle JS, Aspinall EJ, Hutchinson SJ, Quinn B, Gore C, Wiktor SZ, et al. Global policy and access to new hepatitis C therapies for people who inject drugs. The International journal on drug policy. 2015;26(11):1064-71. doi: 10.1016/j.drugpo.2015.05.008

7.   Kamarulzaman A, Verster A, Altice FL. Prisons: ignore them at our peril. Current opinion in HIV and AIDS. 2019;14(5):415-22. doi: 10.1097/coh.0000000000000572

8.   Sagnelli E, Alessio L, Sagnelli C, Gualdieri L, Pisaturo M, Minichini C, et al. Clinical Findings of HCV Chronic Infection in Undocumented Immigrants and Low-Income Refugees in Three Areas of Southern Italy. Annals of hepatology. 2018;17(1):47-53. doi: 10.5604/01.3001.0010.7534

9.   Solomon SS, Solomon S, McFall AM, Srikrishnan AK, Anand S, Verma V, et al. Integrated HIV testing, prevention, and treatment intervention for key populations in India: a cluster-randomised trial. The lancet HIV. 2019;6(5):e283-e96. doi: 10.1016/s2352-3018(19)30034-7

10.   Schweitzer A, Horn J, Mikolajczyk RT, Krause G, Ott JJ. Estimations of worldwide prevalence of chronic hepatitis B virus infection: a systematic review of data published between 1965 and 2013. Lancet (London, England). 2015;386(10003):1546-55. doi: 10.1016/s0140-6736(15)61412-x

11.   Gilmartin CE, Daley AJ, Leung L. The hepatitis B birth-dose immunisation: Exploring parental refusal. The Australian & New Zealand journal of obstetrics & gynaecology. 2020;60(1):93-100. doi: 10.1111/ajo.13008

12.   Okenwa UJ, Dairo MD, Uba B, Ajumobi O. Maternal reasons for non-receipt of valid Hepatitis B birth dose among mother-infant pairs attending routine immunization clinics, South-east, Nigeria. Vaccine. 2019;37(46):6894-9. doi: 10.1016/j.vaccine.2019.09.056

13.   Bassoum O, Kimura M, Tal Dia A, Lemoine M, Shimakawa Y. Coverage and Timeliness of Birth Dose Vaccination in Sub-Saharan Africa: A Systematic Review and

References

Meta-Analysis. Vaccines. 2020;8(2). doi: 10.3390/vaccines8020301

14. Hayashi T, Hutin YJ, Bulterys M, Altaf A, Allegranzi B. Injection practices in 2011-2015: a review using data from the demographic and health surveys (DHS). BMC Health Serv Res. 2019;19(1):600. doi: 10.1186/s12913-019-4366-9

15. Apers L, Vanden Berghe W, De Wit S, Kabeya K, Callens S, Buyze J, et al. Risk factors for HCV acquisition among HIV-positive MSM in Belgium. J Acquir Immune Defic Syndr. 2015;68(5):585-93. doi: 10.1097/qai.0000000000000528

16. Visseaux B, Hué S, Le Hingrat Q, Salmona M, Lebourgeois S, Delaugerre C, et al. Phylogenetic investigation of HCV-4d epidemic in Paris MSM HIV population reveals a still active outbreak and a strong link to the Netherlands. Clin Microbiol Infect. 2020;26(6):785.e1-.e4. doi: 10.1016/j.cmi.2020.01.034

17. Trooskin SB, Reynolds H, Kostman JR. Access to Costly New Hepatitis C Drugs: Medicine, Money, and Advocacy. Clinical infectious diseases : an official publication of the Infectious Diseases Society of America. 2015;61(12):1825-30. doi: 10.1093/cid/civ677

18. Omran D, Alboraie M, Zayed RA, Wifi MN, Naguib M, Eltabbakh M, et al. Towards hepatitis C virus elimination: Egyptian experience, achievements and limitations. World journal of gastroenterology. 2018;24(38):4330-40. doi: 10.3748/wjg.v24.i38.4330

19. Gupta V, Kumar A, Sharma P, Arora A. Newer direct-acting antivirals for hepatitis C virus infection: Perspectives for India. The Indian journal of medical research. 2017;146(1):23-33. doi: 10.4103/ijmr.IJMR_679_15

20. Ward JW, Hinman AR. What Is Needed to Eliminate Hepatitis B Virus and Hepatitis C Virus as Global Health Threats. Gastroenterology. 2019;156(2):297-310. doi: 10.1053/j.gastro.2018.10.048

21. Prevention of hepatitis B and C in the EU/EEA and the UK. Published by: ECDC; 2020. Available from: https://www.ecdc.europa.eu/sites/default/files/documents/hepatitis-B-and-C-prevention_1.pdf

22. Duffell E, Noori T. Monitoring the responses to hepatitis B and C epidemics in the EU/EEA Member States. Published by: ECDC; 2020. Available from: https://www.ecdc.europa.eu/en/publications-data/monitoring-responses-hepatitis-b-and-c-epidemics-eueea-member-states-2019

23. Hepatitis B and C testing in the EU/EEA: progress in reaching the elimination targets. . Published by: ECDC; 2021. Available from: https://www.ecdc.europa.eu/en/publications-data/hepatitis-b-and-c-testing-eueea-progress-reaching-elimination-targets

24. The sustainable development goals and hepatitis B and C in the EU/EEA. Published by: ECDC; 2021. Available from: https://www.ecdc.europa.eu/en/publications-data/hepatitis-b-and-c-sustainable-development-goals-eu-eea

25. Monitoring the elimination of viral hepatitis as a public health threat among people who inject drugs in Europe: the elimination barometer. Published by: EMCDDA; 2019. Available from:

https://www.emcdda.europa.eu/system/files/publications/11796/Technical%20report_The%20elimination%20barometer%20for%20viral%20hepatitis%20among%20PWID%20in%20Europe_0.pdf

26.   Monitoring and evaluation for viral hepatitis B and C: recommended indicators and framework. Published by: WHO; 2016. Available from: https://www.who.int/publications/i/item/monitoring-and-evaluation-for-viral-hepatitis-b-and-c-recommended-indicators-and-framework

27.   Mårdh O, Quinten C, Amato-Gauci AJ, Duffell E. Mortality from liver diseases attributable to hepatitis B and C in the EU/EEA - descriptive analysis and estimation of 2015 baseline. Infectious diseases (London, England). 2020;52(9):625-37. doi: 10.1080/23744235.2020.1766104

28.   Carvalhana SC, Leitão J, Alves AC, Bourbon M, Cortez-Pinto H. Hepatitis B and C prevalence in Portugal: disparity between the general population and high-risk groups. European journal of gastroenterology & hepatology. 2016;28(6):640-4. doi: 10.1097/meg.0000000000000608

29.   Inquérito Serológico Nacional 2015-2016. Published by: INSA; 2017. Report No.: 978-989-8794-42-0. Available from: https://repositorio.insa.pt/bitstream/10400.18/5402/1/INSA_ISN-2015-2016-IST_web.pdf

30.   Anjo J, Café A, Carvalho A, Doroana M, Fraga J, Gíria J, et al. O impacto da hepatite C em Portugal. Jornal Português de Gastrenterologia. 2014;21:44-54. doi: 10.1016/j.jpg.2014.03.001

31.   Marinho RT, Moura MC, Giria JA, Ferrinho P. Epidemiological aspects of hepatitis C in Portugal. Journal of gastroenterology and hepatology. 2001;16(9):1076-7. doi: 10.1046/j.1440-1746.2001.2574a.x

32.   Relatório Anual - A Situação do País em Matéria de Drogas e Toxicodependências. Published by: SICAD; 2019. Available from: http://www.sicad.pt/BK/Publicacoes/Lists/SICAD_PUBLICACOES/Attachments/162/Relatorio_Anual_2018_%20ANEXO_A_SituacaoDoPaisEmMateriaDeDrogas_e_Toxicodependencias.pdf

33.   Gaspar R, Liberal R, Esteves J, Morgado R, Macedo G. Prison Break: Curing HCV Infection in the Prison Setting. Official journal of the American College of Gastroenterology | ACG. 2018;113:S542. doi: 10.14309/00000434-201810001-00963

34.   Norma da Direção Geral da Saúde - Tratamento Hepatite C. Published by: DGS; 2015. Available from: https://www.pnvihsida.dgs.pt/informacao-tecnica-e-cientifica111/legislacao-em-vigor/norma-0112012-de-16-de-dezembro-atualizada-em-30-de-abril-de-2015-pdf.aspx

35.   Programa Nacional Para as Hepatites Virais da Direção Geral de Saúde. Published by: DGS; 2017. Available from: https://www.hepb.org/assets/Uploads/DGS-PNHV2017-V7-National-PLan-Portugal-2017.pdf

References

36.   Hepatite C – Monitorização dos tratamentos. Published by: INFARMED; 2019. Available from: https://www.infarmed.pt/documents/15786/1084985/Hepatite+C/dd545649-1042-42e9-b2d3-83db2bc34b11

37.   INFARMED. Hepatite C – Monitorização dos tratamentos. Published by: INFARMED;   2021.   Available   from: https://www.infarmed.pt/documents/15786/1084985/Hepatite+C/dd545649-1042-42e9-b2d3-83db2bc34b11

38.   Programa Nacional Para as Hepatites Virais da Direção Geral de Saúde. Published by: DGS; 2019. Available from: https://www.dgs.pt/portal-da-estatistica-da-saude/diretorio-de-informacao/diretorio-de-informacao/por-serie-1116211-pdf.aspx?v=%3d%3dDwAAAB%2bLCAAAAAAABAArySzItzVUy81MsTU1MDAFAHzFEfkPAAAA

39.   Silva MJ, Valente J, Capela T, Russo P, Calinas F. Epidemiology of hepatitis B in Portugal. European journal of gastroenterology & hepatology. 2017;29(3):249-58. doi: 10.1097/meg.0000000000000792

40.   Lecour H, Ribeiro AT, Amaral I, Rodrigues MA. Prevalence of viral hepatitis markers in the population of Portugal. Bulletin of the World Health Organization. 1984;62(5):743-7.

41.   Santos A, Carvalho A, Tomaz J, Rodrigues V, Coxinho L, Bento D, et al. Prevalence of markers of hepatitis B infection in the adult population of the district of Coimbra. Acta Médica Portuguesa. 2000;13(4):167-71. doi: 10.20344/amp.1790

42.   Silva MJ, Rosa MV, Nogueira PJ, Calinas F. Ten years of hospital admissions for liver cirrhosis in Portugal. European journal of gastroenterology & hepatology. 2015;27(11):1320-6. doi: 10.1097/meg.0000000000000449

43.   Epidemiological assessment of hepatitis B and C among migrants in the EU/EEA. Published   by: ECDC;   2016.   Available   from: https://www.ecdc.europa.eu/sites/default/files/media/en/publications/Publications/epidemiological-assessment-hepatitis-B-and-C-among-migrants-EU-EEA.pdf

44.   Falcão F, Lopes C, Viegas E, Perez R, Aldir I, Farinha H, et al. Experience of a Portuguese Center: Effectiveness of Direct-Acting Antiviral Therapy for Hepatitis C. Acta medica portuguesa. 2019;32(3):189-94. doi: 10.20344/amp.10655

45.   Simmonds P, Becher P, Bukh J, Gould EA, Meyers G, Monath T, et al. ICTV Virus Taxonomy Profile: Flaviviridae. The Journal of general virology. 2017;98(1):2-3. doi: 10.1099/jgv.0.000672

46.   Li HC, Yang CH, Lo SY. Hepatitis C Viral Replication Complex. Viruses. 2021;13(3). doi: 10.3390/v13030520

47.   Ganta NM, Gedda G, Rathnakar B, Satyanarayana M, Yamajala B, Ahsan MJ, et al. A review on HCV inhibitors: Significance of non-structural polyproteins. European journal of medicinal chemistry. 2019;164:576-601. doi: 10.1016/j.ejmech.2018.12.045

48.   Sidorkiewicz M. Hepatitis C Virus Uses Host Lipids to Its Own Advantage.

Metabolites. 2021;11(5). doi: 10.3390/metabo11050273

49.   Manns MP, Buti M, Gane E, Pawlotsky JM, Razavi H, Terrault N, et al. Hepatitis C virus infection. Nat Rev Dis Primers. 2017;3:17006. doi: 10.1038/nrdp.2017.6

50.   Herker E, Ott M. Unique ties between hepatitis C virus replication and intracellular lipids. Trends Endocrinol Metab. 2011;22(6):241-8. doi: 10.1016/j.tem.2011.03.004

51.   Magnius L, Mason WS, Taylor J, Kann M, Glebe D, Dény P, et al. ICTV Virus Taxonomy Profile: Hepadnaviridae. The Journal of general virology. 2020;101(6):571-2. doi: 10.1099/jgv.0.001415

52.   McNaughton AL, D'Arienzo V, Ansari MA, Lumley SF, Littlejohn M, Revill P, et al. Insights From Deep Sequencing of the HBV Genome-Unique, Tiny, and Misunderstood. Gastroenterology. 2019;156(2):384-99. doi: 10.1053/j.gastro.2018.07.058

53.   Tsukuda S, Watashi K. Hepatitis B virus biology and life cycle. Antiviral Res. 2020;182:104925. doi: 10.1016/j.antiviral.2020.104925

54.   Prifti GM, Moianos D, Giannakopoulou E, Pardali V, Tavis JE, Zoidis G. Recent Advances in Hepatitis B Treatment. Pharmaceuticals (Basel, Switzerland). 2021;14(5). doi: 10.3390/ph14050417

55.   Valaydon ZS, Locarnini SA. The virological aspects of hepatitis B. Best Pract Res Clin Gastroenterol. 2017;31(3):257-64. doi: 10.1016/j.bpg.2017.04.013

56.   Caballero A, Tabernero D, Buti M, Rodriguez-Frias F. Hepatitis B virus: The challenge of an ancient virus with multiple faces and a remarkable replication strategy. Antiviral Res. 2018;158:34-44. doi: 10.1016/j.antiviral.2018.07.019

57.   Lucifora J, Arzberger S, Durantel D, Belloni L, Strubin M, Levrero M, et al. Hepatitis B virus X protein is essential to initiate and maintain virus replication after infection. J Hepatol. 2011;55(5):996-1003. doi: 10.1016/j.jhep.2011.02.015

58.   Wei Y, Neuveut C, Tiollais P, Buendia MA. Molecular biology of the hepatitis B virus and role of the X gene. Pathologie-biologie. 2010;58(4):267-72. doi: 10.1016/j.patbio.2010.03.005

59.   Moriishi K, Matsuura Y. Exploitation of lipid components by viral and host proteins for hepatitis C virus infection. Front Microbiol. 2012;3:54. doi: 10.3389/fmicb.2012.00054

60.   Wang J, Huang H, Liu Y, Chen R, Yan Y, Shi S, et al. Hepatitis B Virus Infection - Molecular Virology to Antiviral Drugs. Chapter 2 - HBV Genome and Life Cycle. Series: Advances in Experimental Medicine and Biology. Vol. 1179. Published by Springer; 2020. p. 17-38.

61.   Park SG, Kim Y, Park E, Ryu HM, Jung G. Fidelity of hepatitis B virus polymerase. European journal of biochemistry. 2003;270(14):2929-36. doi: 10.1046/j.1432-1033.2003.03650.x

62.   Echeverría N, Moratorio G, Cristina J, Moreno P. Hepatitis C virus genetic variability and evolution. World journal of hepatology. 2015;7(6):831-45. doi: 10.4254/wjh.v7.i6.831

63.   Revill PA, Tu T, Netter HJ, Yuen LKW, Locarnini SA, Littlejohn M. The evolution and clinical impact of hepatitis B virus genome diversity. Nat Rev Gastroenterol Hepatol. 2020;17(10):618-34. doi: 10.1038/s41575-020-0296-6

64.   Pourkarim MR, Amini-Bavil-Olyaee S, Lemey P, Maes P, Van Ranst M. Are hepatitis B virus "subgenotypes" defined accurately? J Clin Virol. 2010;47(4):356-60. doi: 10.1016/j.jcv.2010.01.015

65.   Mojsiejczuk LN, Torres C, Fainboin HA, Galdame OA, Campos RH, Flichman DM. Identification of a new clade of hepatitis B virus genotype F. Infection, genetics and evolution : journal of molecular epidemiology and evolutionary genetics in infectious diseases. 2015;34:122-5. doi: 10.1016/j.meegid.2015.06.007

66.   Wolf JM, Pereira V, Simon D, Lunge VR. Evolutionary history of hepatitis B virus genotype H. Journal of medical virology. 2021;93(6):4004-9. doi: 10.1002/jmv.26463

67.   Kramvis A. Genotypes and genetic variability of hepatitis B virus. Intervirology. 2014;57(3-4):141-50. doi: 10.1159/000360947

68.   Pourkarim MR, Amini-Bavil-Olyaee S, Kurbanov F, Van Ranst M, Tacke F. Molecular identification of hepatitis B virus genotypes/subgenotypes: revised classification hurdles and updated resolutions. World journal of gastroenterology. 2014;20(23):7152-68. doi: 10.3748/wjg.v20.i23.7152

69.   McNaughton AL, Revill PA, Littlejohn M, Matthews PC, Ansari MA. Analysis of genomic-length HBV sequences to determine genotype and subgenotype reference sequences. The Journal of general virology. 2020;101(3):271-83. doi: 10.1099/jgv.0.001387

70.   Bukh J. The history of hepatitis C virus (HCV): Basic research reveals unique features in phylogeny, evolution and the viral life cycle with new perspectives for epidemic control. J Hepatol. 2016;65(1 Suppl):S2-s21. doi: 10.1016/j.jhep.2016.07.035

71.   Smith DB, Bukh J, Kuiken C, Muerhoff AS, Rice CM, Stapleton JT, et al. Expanded classification of hepatitis C virus into 7 genotypes and 67 subtypes: updated criteria and genotype assignment web resource. Hepatology (Baltimore, Md). 2014;59(1):318-27. doi: 10.1002/hep.26744

72.   Borgia SM, Hedskog C, Parhy B, Hyland RH, Stamm LM, Brainard DM, et al. Identification of a Novel Hepatitis C Virus Genotype From Punjab, India: Expanding Classification of Hepatitis C Virus Into 8 Genotypes. The Journal of infectious diseases. 2018;218(11):1722-9. doi: 10.1093/infdis/jiy401

73.   Pickett BE, Striker R, Lefkowitz EJ. Evidence for separation of HCV subtype 1a into two distinct clades. Journal of viral hepatitis. 2011;18(9):608-18. doi: 10.1111/j.1365-2893.2010.01342.x

74.   Smith DB, Bukh J, Kuiken C, Muerhoff AS, Rice CM, Stapleton JT, et al. Confirmed HCV genotypes/subtypes. International Committee on Taxonomy of Viruses (ICTV); 2019. updated May, 2019; Available from: https://talk.ictvonline.org/ictv_wikis/flaviviridae/w/sg_flavi/56/hcv-classification.

75.   Messina JP, Humphreys I, Flaxman A, Brown A, Cooke GS, Pybus OG, et al. Global distribution and prevalence of hepatitis C virus genotypes. Hepatology (Baltimore, Md). 2015;61(1):77-87. doi: 10.1002/hep.27259

76.   Kalinina O, Norder H, Mukomolov S, Magnius LO. A natural intergenotypic recombinant of hepatitis C virus identified in St. Petersburg. J Virol. 2002;76(8):4034-43. doi: 10.1128/jvi.76.8.4034-4043.2002

77.   Tatematsu K, Tanaka Y, Kurbanov F, Sugauchi F, Mano S, Maeshiro T, et al. A genetic variant of hepatitis B virus divergent from known human and ape genotypes isolated from a Japanese patient and provisionally assigned to new genotype J. J Virol. 2009;83(20):10538-47. doi: 10.1128/JVI.00462-09

78.   Araujo NM, Teles SA, Spitz N. Comprehensive Analysis of Clinically Significant Hepatitis B Virus Mutations in Relation to Genotype, Subgenotype and Geographic Region. Front Microbiol. 2020;11:616023. doi: 10.3389/fmicb.2020.616023

79.   Tran TT, Trinh TN, Abe K. New complex recombinant genotype of hepatitis B virus identified in Vietnam. J Virol. 2008;82(11):5657-63. doi: 10.1128/jvi.02556-07

80.   Murphy DG, Sablon E, Chamberland J, Fournier E, Dandavino R, Tremblay CL. Hepatitis C virus genotype 7, a new genotype originating from central Africa. J Clin Microbiol. 2015;53(3):967-72. doi: 10.1128/jcm.02831-14

81.   Salmona M, Caporossi A, Simmonds P, Thélu MA, Fusillier K, Mercier-Delarue S, et al. First next-generation sequencing full-genome characterization of a hepatitis C virus genotype 7 divergent subtype. Clin Microbiol Infect. 2016;22(11):947.e1-.e8. doi: 10.1016/j.cmi.2016.07.032

82.   Jackowiak P, Kuls K, Budzko L, Mania A, Figlerowicz M, Figlerowicz M. Phylogeny and molecular evolution of the hepatitis C virus. Infection, genetics and evolution : journal of molecular epidemiology and evolutionary genetics in infectious diseases. 2014;21:67-82. doi: 10.1016/j.meegid.2013.10.021

83.   Frank C, Mohamed MK, Strickland GT, Lavanchy D, Arthur RR, Magder LS, et al. The role of parenteral antischistosomal therapy in the spread of hepatitis C virus in Egypt. Lancet (London, England). 2000;355(9207):887-91. doi: 10.1016/s0140-6736(99)06527-7

84.   Esteban JI, Sauleda S, Quer J. The changing epidemiology of hepatitis C virus infection in Europe. J Hepatol. 2008;48(1):148-62. doi: 10.1016/j.jhep.2007.07.033

85.   de Bruijne J, Schinkel J, Prins M, Koekkoek SM, Aronson SJ, van Ballegooijen MW, et al. Emergence of hepatitis C virus genotype 4: phylogenetic analysis reveals three distinct epidemiological profiles. J Clin Microbiol. 2009;47(12):3832-8. doi: 10.1128/JCM.01146-09

86.   Fu Y, Qin W, Cao H, Xu R, Tan Y, Lu T, et al. HCV 6a prevalence in Guangdong province had the origin from Vietnam and recent dissemination to other regions of China: phylogeographic analyses. PloS one. 2012;7(1):e28006. doi: 10.1371/journal.pone.0028006

87.   Verbeeck J, Maes P, Lemey P, Pybus OG, Wollants E, Song E, et al. Investigating the origin and spread of hepatitis C virus genotype 5a. J Virol. 2006;80(9):4220-6. doi:

References

10.1128/JVI.80.9.4220-4226.2006

88.   Velkov S, Ott JJ, Protzer U, Michler T. The Global Hepatitis B Virus Genotype Distribution Approximated from Available Genotyping Data. Genes (Basel). 2018;9(10). doi: 10.3390/genes9100495

89.   Sunbul M. Hepatitis B virus genotypes: global distribution and clinical importance. World journal of gastroenterology. 2014;20(18):5427-34. doi: 10.3748/wjg.v20.i18.5427

90.   Kostaki EG, Karamitros T, Stefanou G, Mamais I, Angelis K, Hatzakis A, et al. Unravelling the history of hepatitis B virus genotypes A and D infection using a full-genome phylogenetic and phylogeographic approach. eLife. 2018;7. doi: 10.7554/eLife.36709

91.   Spitz N, Mello FCA, Moreira AS, Gusatti CS, Martins RMB, Gomes SA, et al. Reconstruction of the spatial and temporal dynamics of hepatitis B virus genotype D in the Americas. PloS one. 2019;14(7):e0220342. doi: 10.1371/journal.pone.0220342

92.   Osiowy C, Larke B, Giles E. Distinct geographical and demographic distribution of hepatitis B virus genotypes in the Canadian Arctic as revealed through an extensive molecular epidemiological survey. Journal of viral hepatitis. 2011;18(4):e11-9. doi: 10.1111/j.1365-2893.2010.01356.x

93.   Bouckaert R, Simons BC, Krarup H, Friesen TM, Osiowy C. Tracing hepatitis B virus (HBV) genotype B5 (formerly B6) evolutionary history in the circumpolar Arctic through phylogeographic modelling. PeerJ. 2017;5:e3757. doi: 10.7717/peerj.3757

94.   Andernach IE, Hunewald OE, Muller CP. Bayesian inference of the evolution of HBV/E. PloS one. 2013;8(11):e81690. doi: 10.1371/journal.pone.0081690

95.   Stuyver L, De Gendt S, Van Geyt C, Zoulim F, Fried M, Schinazi RF, et al. A new genotype of hepatitis B virus: complete genome and phylogenetic relatedness. The Journal of general virology. 2000;81(Pt 1):67-74. doi: 10.1099/0022-1317-81-1-67

96.   Wolf JM, De Carli S, Pereira V, Simon D, Lunge VR. Temporal evolution and global spread of hepatitis B virus genotype G. Journal of viral hepatitis. 2021;28(2):393-9. doi: 10.1111/jvh.13431

97.   Pourkarim MR, Lemey P, Amini-Bavil-Olyaee S, Houspie L, Verbeeck J, Rahman M, et al. Molecular characterization of hepatitis B virus strains circulating in Belgian patients co-infected with HIV and HBV: overt and occult infection. Journal of medical virology. 2011;83(11):1876-84. doi: 10.1002/jmv.22174

98.   Cornelissen M, Zorgdrager F, Bruisten SM, Bakker M, Berkhout B, van der Kuyl AC. Widespread hepatitis B virus genotype G (HBV-G) infection during the early years of the HIV epidemic in the Netherlands among men who have sex with men. BMC infectious diseases. 2016;16:268. doi: 10.1186/s12879-016-1599-7

99.   Basic M, Kubesch A, Kuhnhenn L, Görgülü E, Finkelmeier F, Dietz J, et al. Not uncommon: HBV genotype G co-infections among healthy European HBV carriers with genotype A and E infection. Liver Int. 2021;41(6):1278-89. doi: 10.1111/liv.14884

100.   Rodrigues A, Pinho L, Lobato L, Queirós A, Castro R, Daniel M, et al. Hepatitis C

virus genotypes and the influence of the induction of immunosuppression with anti-thymocyte globulin (ATG) on chronic hepatitis in renal graft recipients. Transplant international : official journal of the European Society for Organ Transplantation. 1998;11 Suppl 1:S115-8. doi: 10.1007/s001470050441

101.   Ramalho F, Costa A, Pires A, Cabrita P, Serejo F, Correia AP, et al. Correlation of genotypes and route of transmission with histologic activity and disease stage in chronic hepatitis C. Digestive diseases and sciences. 2000;45(1):182-7. doi: 10.1023/a:1005442317680

102.   Sarmento-Castro R, Horta A, Vasconcelos O, Coelho H, Mendez J, Tavares AP, et al. Impact of peginterferon alpha-2b and ribavirin treatment on liver tissue in patients with HCV or HCV-HIV co-infection. The Journal of infection. 2007;54(6):609-16. doi: 10.1016/j.jinf.2006.11.010

103.   Calado RA, Rocha MR, Parreira R, Piedade J, Venenno T, Esteves A. Hepatitis C virus subtypes circulating among intravenous drug users in Lisbon, Portugal. Journal of medical virology. 2011;83(4):608-15. doi: 10.1002/jmv.21955

104.   Pádua E, Avó AP, Almeida C, Água Doce I, Cortes Martins H. Assessment of Hepatitis C Virus Diversity in Addition to the Frequency of Genotypes in Samples Analyzed Between 2009 and 2014 at the Reference Laboratory of National Health Institute Dr. Ricardo Jorge. Acta medica portuguesa. 2015;28(6):695-701.

105.   Mota A, Guedes F, Areias J, Pinho L, Cardoso MF. Epidemiological study of genotypes of hepatitis B virus in northern Portugal. Journal of medical virology. 2009;81(7):1170-6. doi: 10.1002/jmv.21526

106.   Marcelino R, Mimoso P, Matos S, Velosa J, Glória HRM. Genetic diversity of hepatitis B virus and resistance mutations in Portugal. 11th Meeting of the Associação Portuguesa para o Estudo do Fígado; 2008.

107.   Geoghegan JL, Holmes EC. Evolutionary Virology at 40. Genetics. 2018;210(4):1151-62. doi: 10.1534/genetics.118.301556

108.   Beamud B, Bracho MA, González-Candelas F. Characterization of New Recombinant Forms of HIV-1 From the Comunitat Valenciana (Spain) by Phylogenetic Incongruence. Front Microbiol. 2019;10:1006. doi: 10.3389/fmicb.2019.01006

109.   Pyrc K, Dijkman R, Deng L, Jebbink MF, Ross HA, Berkhout B, et al. Mosaic structure of human coronavirus NL63, one thousand years of evolution. Journal of molecular biology. 2006;364(5):964-73. doi: 10.1016/j.jmb.2006.09.074

110.   Miao J, Ran J, Song Y, Liu Y, Gao L, Miao Z, et al. Characterization of a Novel HIV-1 Circulating Recombinant Form, CRF01_AE/B'/C (CRF96_cpx), in Yunnan, China. AIDS research and human retroviruses. 2018;34(4):393-7. doi: 10.1089/aid.2017.0288

111.   Lemey P, Salemi M, Vandamme A. The Phylogenetic Handbook: A Practical Approach to Phylogenetic Analysis and Hypothesis Testing. 2nd ed: Cambridge University Press; 2009.

References

112.   Dudas G, Carvalho LM, Rambaut A, Bedford T. MERS-CoV spillover at the camel-human interface. eLife. 2018;7. doi: 10.7554/eLife.31257

113.   Dudas G, Carvalho LM, Bedford T, Tatem AJ, Baele G, Faria NR, et al. Virus genomes reveal factors that spread and sustained the Ebola epidemic. Nature. 2017;544(7650):309-15. doi: 10.1038/nature22040

114.   Faria NR, Quick J, Claro IM, Theze J, de Jesus JG, Giovanetti M, et al. Establishment and cryptic transmission of Zika virus in Brazil and the Americas. Nature. 2017;546(7658):406-10. doi: 10.1038/nature22401

115.   Bedford T, Suchard MA, Lemey P, Dudas G, Gregory V, Hay AJ, et al. Integrating influenza antigenic dynamics with molecular evolution. eLife. 2014;3:e01914. doi: 10.7554/eLife.01914

116.   Neher RA, Bedford T. nextflu: real-time tracking of seasonal influenza virus evolution in humans. Bioinformatics. 2015;31(21):3546-8. doi: 10.1093/bioinformatics/btv381

117.   Cui H, Shi Y, Ruan T, Li X, Teng Q, Chen H, et al. Phylogenetic analysis and pathogenicity of H3 subtype avian influenza viruses isolated from live poultry markets in China. Sci Rep. 2016;6:27360. doi: 10.1038/srep27360

118.   Sallard E, Halloy J, Casane D, Decroly E, van Helden J. Tracing the origins of SARS-COV-2 in coronavirus phylogenies: a review. Environ Chem Lett. 2021:1-17. doi: 10.1007/s10311-020-01151-1

119.   Geoghegan JL, Senior AM, Di Giallonardo F, Holmes EC. Virological factors that increase the transmissibility of emerging human viruses. Proc Natl Acad Sci U S A. 2016;113(15):4170-5. doi: 10.1073/pnas.1521582113

120.   Geoghegan JL, Duchene S, Holmes EC. Comparative analysis estimates the relative frequencies of co-divergence and cross-species transmission within viral families. PLoS Pathog. 2017;13(2):e1006215. doi: 10.1371/journal.ppat.1006215

121.   Rife BD, Mavian C, Chen X, Ciccozzi M, Salemi M, Min J, et al. Phylodynamic applications in 21(st) century global infectious disease research. Glob Health Res Policy. 2017;2:13. doi: 10.1186/s41256-017-0034-y

122.   Faria NR, Rambaut A, Suchard MA, Baele G, Bedford T, Ward MJ, et al. HIV epidemiology. The early spread and epidemic ignition of HIV-1 in human populations. Science. 2014;346(6205):56-61. doi: 10.1126/science.1256739

123.   Gray RR, Tatem AJ, Lamers S, Hou W, Laeyendecker O, Serwadda D, et al. Spatial phylodynamics of HIV-1 epidemic emergence in east Africa. AIDS. 2009;23(14):F9-F17. doi: 10.1097/QAD.0b013e32832faf61

124.   Faria NR, Suchard MA, Abecasis A, Sousa JD, Ndembi N, Bonfim I, et al. Phylodynamics of the HIV-1 CRF02_AG clade in Cameroon. Infection, genetics and evolution : journal of molecular epidemiology and evolutionary genetics in infectious diseases. 2012;12(2):453-60. doi: 10.1016/j.meegid.2011.04.028

125.   Faria NR, Suchard MA, Rambaut A, Lemey P. Toward a quantitative understanding of viral phylogeography. Curr Opin Virol. 2011;1(5):423-9. doi: 10.1016/j.coviro.2011.10.003

126.   Theys K, Lemey P, Vandamme AM, Baele G. Advances in Visualization Tools for Phylogenomic and Phylodynamic Studies of Viral Diseases. Front Public Health. 2019;7:208. doi: 10.3389/fpubh.2019.00208

127.   Sanjuán R, Domingo-Calap P. Mechanisms of viral mutation. Cellular and Molecular Life Sciences. 2016;73(23):4433-48. doi: 10.1007/s00018-016-2299-6

128.   Dolan PT, Whitfield ZJ, Andino R. Mechanisms and Concepts in RNA Virus Population Dynamics and Evolution. Annu Rev Virol. 2018;5(1):69-92. doi: 10.1146/annurev-virology-101416-041718

129.   Vrancken B, Suchard MA, Lemey P. Accurate quantification of within- and between-host HBV evolutionary rates requires explicit transmission chain modelling. Virus evolution. 2017;3(2):vex028. doi: 10.1093/ve/vex028

130.   Holder M, Lewis PO. Phylogeny estimation: traditional and Bayesian approaches. Nature reviews Genetics. 2003;4(4):275-84. doi: 10.1038/nrg1044

131.   Yang Z, Rannala B. Molecular phylogenetics: principles and practice. Nature reviews Genetics. 2012;13(5):303-14. doi: 10.1038/nrg3186

132.   Pybus OG, Rambaut A. Evolutionary analysis of the dynamics of viral infectious disease. Nature reviews Genetics. 2009;10(8):540-50. doi: 10.1038/nrg2583

133.   Holmes EC. The phylogeography of human viruses. Mol Ecol. 2004;13(4):745-56. doi: 10.1046/j.1365-294x.2003.02051.x

134.   Pybus OG, Drummond AJ, Nakano T, Robertson BH, Rambaut A. The epidemiology and iatrogenic transmission of hepatitis C virus in Egypt: a Bayesian coalescent approach. Molecular biology and evolution. 2003;20(3):381-7. doi: 10.1093/molbev/msg043

135.   Souza BF, Drexler JF, Lima RS, Rosário Mde O, Netto EM. Theories about evolutionary origins of human hepatitis B virus in primates and humans. The Brazilian journal of infectious diseases : an official publication of the Brazilian Society of Infectious Diseases. 2014;18(5):535-43. doi: 10.1016/j.bjid.2013.12.006

136.   Lin SY, Toyoda H, Kumada T, Liu HF. Molecular Evolution and Phylodynamics of Acute Hepatitis B Virus in Japan. PloS one. 2016;11(6):e0157103. doi: 10.1371/journal.pone.0157103

137.   Littlejohn M, Locarnini S, Yuen L. Origins and Evolution of Hepatitis B Virus and Hepatitis D Virus. Cold Spring Harbor perspectives in medicine. 2016;6(1):a021360. doi: 10.1101/cshperspect.a021360

138.   Magnius LO, Norder H. Subtypes, genotypes and molecular epidemiology of the hepatitis B virus as reflected by sequence variability of the S-gene. Intervirology. 1995;38(1-2):24-34. doi: 10.1159/000150411

139.   Paraskevis D, Magiorkinis G, Magiorkinis E, Ho SY, Belshaw R, Allain JP, et al. Dating the origin and dispersal of hepatitis B virus infection in humans and primates. Hepatology (Baltimore, Md). 2013;57(3):908-16. doi: 10.1002/hep.26079

140.   Bollyky P RA, Grassley N, Carman W and Holmes E, editor Hepatitis B virus has a recent new world evolutionary origin (Poster n.º P/C05/009). American Association for the Study of Liver Diseases Conference; 1997: Published in Journal of Hepatology

141.   Kahila Bar-Gal G, Kim MJ, Klein A, Shin DH, Oh CS, Kim JW, et al. Tracing hepatitis B virus to the 16th century in a Korean mummy. Hepatology (Baltimore, Md). 2012;56(5):1671-80. doi: 10.1002/hep.25852

142.   MacDonald DM, Holmes EC, Lewis JC, Simmonds P. Detection of hepatitis B virus infection in wild-born chimpanzees (Pan troglodytes verus): phylogenetic relationships with human and other primate genotypes. J Virol. 2000;74(9):4253-7. doi: 10.1128/jvi.74.9.4253-4257.2000

143.   Krause-Kyora B, Susat J, Key FM, Kühnert D, Bosse E, Immel A, et al. Neolithic and medieval virus genomes reveal complex evolution of hepatitis B. eLife. 2018;7. doi: 10.7554/eLife.36666

144.   Mühlemann B, Jones TC, Damgaard PB, Allentoft ME, Shevnina I, Logvin A, et al. Ancient hepatitis B viruses from the Bronze Age to the Medieval period. Nature. 2018;557(7705):418-23. doi: 10.1038/s41586-018-0097-z

145.   Datta S. Excavating new facts from ancient Hepatitis B virus sequences. Virology. 2020;549:89-99. doi: 10.1016/j.virol.2020.08.002

146.   Zehender G, Svicher V, Gabanelli E, Ebranati E, Veo C, Lo Presti A, et al. Reliable timescale inference of HBV genotype A origin and phylodynamics. Infection, genetics and evolution : journal of molecular epidemiology and evolutionary genetics in infectious diseases. 2015;32:361-9. doi: 10.1016/j.meegid.2015.03.009

147.   Kramvis A, Paraskevis D. Subgenotype A1 of HBV--tracing human migrations in and out of Africa. Antiviral therapy. 2013;18(3 Pt B):513-21. doi: 10.3851/imp2657

148.   Quintero A, Martínez D, Alarcón De Noya B, Costagliola A, Urbina L, González N, et al. Molecular epidemiology of hepatitis B virus in Afro-Venezuelan populations. Archives of virology. 2002;147(9):1829-36. doi: 10.1007/s00705-002-0842-2

149.   Motta-Castro AR, Martins RM, Yoshida CF, Teles SA, Paniago AM, Lima KM, et al. Hepatitis B virus infection in isolated Afro-Brazilian communities. Journal of medical virology. 2005;77(2):188-93. doi: 10.1002/jmv.20435

150.   Gusatti CS, Costi C, Halon ML, Grandi T, Medeiros AF, Silva CM, et al. Hepatitis B Virus Genotype D Isolates Circulating in Chapeco, Southern Brazil, Originate from Italy. PloS one. 2015;10(8):e0135816. doi: 10.1371/journal.pone.0135816

151.   Lago BV, do Espirito-Santo MP, Costa VD, Marques VA, Villar LM, Lewis-Ximenez LL, et al. Genetic Diversity of the Hepatitis B Virus Subgenotypes in Brazil. Viruses. 2019;11(9). doi: 10.3390/v11090860

152. Bertolini DA, Gomes-Gouvêa MS, Guedes de Carvalho-Mello IM, Saraceni CP, Sitnik R, Grazziotin FG, et al. Hepatitis B virus genotypes from European origin explains the high endemicity found in some areas from southern Brazil. Infection, genetics and evolution : journal of molecular epidemiology and evolutionary genetics in infectious diseases. 2012;12(6):1295-304. doi: 10.1016/j.meegid.2012.04.009

153. Alonso B. The Other Europeans: Immigration into Latin America and the International Labour Market (1870–1930). Revista de Historia Económica / Journal of Iberian and Latin American Economic History Cambridge University Press. 2007;25(3):395–426. doi: 10.1017/S0212610900000185

154. Mulders MN, Venard V, Njayou M, Edorh AP, Bola Oyefolu AO, Kehinde MO, et al. Low genetic diversity despite hyperendemicity of hepatitis B virus genotype E throughout West Africa. The Journal of infectious diseases. 2004;190(2):400-8. doi: 10.1086/421502

155. Andernach IE, Hübschen JM, Muller CP. Hepatitis B virus: the genotype E puzzle. Reviews in medical virology. 2009;19(4):231-40. doi: 10.1002/rmv.618

156. Lin SYC, Magalis BR, Salemi M, Liu HF. Origin and dissemination of hepatitis B virus genotype C in East Asia revealed by phylodynamic analysis and historical correlates. Journal of viral hepatitis. 2019;26(1):145-54. doi: 10.1111/jvh.13006

157. Mello FCA, Araujo OC, Lago BV, Motta-Castro ARC, Moraes MTB, Gomes SA, et al. Phylogeography and evolutionary history of hepatitis B virus genotype F in Brazil. Virology Journal. 2013;10(1):236. doi: 10.1186/1743-422X-10-236

158. Pybus OG, Gray RR. The virus whose family expanded. Nature. 2013;498(7454):310-1. doi: 10.1038/498310b

159. Pybus OG, Thézé J. Hepacivirus cross-species transmission and the origins of the hepatitis C virus. Curr Opin Virol. 2016;16:1-7. doi: 10.1016/j.coviro.2015.10.002

160. Pfaender S, Brown RJ, Pietschmann T, Steinmann E. Natural reservoirs for homologs of hepatitis C virus. Emerg Microbes Infect. 2014;3(3):e21. doi: 10.1038/emi.2014.19

161. Scheel TK, Simmonds P, Kapoor A. Surveying the global virome: identification and characterization of HCV-related animal hepaciviruses. Antiviral Res. 2015;115:83-93. doi: 10.1016/j.antiviral.2014.12.014

162. Simmonds P. The origin of hepatitis C virus. Curr Top Microbiol Immunol. 2013;369:1-15. doi: 10.1007/978-3-642-27340-7_1

163. Smith DB, Pathirana S, Davidson F, Lawlor E, Power J, Yap PL, et al. The origin of hepatitis C virus genotypes. The Journal of general virology. 1997;78 (Pt 2):321-8. doi: 10.1099/0022-1317-78-2-321

164. Pybus OG, Charleston MA, Gupta S, Rambaut A, Holmes EC, Harvey PH. The epidemic behavior of the hepatitis C virus. Science. 2001;292(5525):2323-5. doi: 10.1126/science.1058321

165. Kapoor A, Simmonds P, Gerold G, Qaisar N, Jain K, Henriquez JA, et al. Characterization of a canine homolog of hepatitis C virus. Proc Natl Acad Sci U S A.

References

2011;108(28):11608-13. doi: 10.1073/pnas.1101794108

166.   Li C, Lu L, Murphy DG, Negro F, Okamoto H. Origin of hepatitis C virus genotype 3 in Africa as estimated through an evolutionary analysis of the full-length genomes of nine subtypes, including the newly sequenced 3d and 3e. The Journal of general virology. 2014;95(Pt 8):1677-88. doi: 10.1099/vir.0.065128-0

167.   Lu L, Li C, Xu Y, Murphy DG. Full-length genomes of 16 hepatitis C virus genotype 1 isolates representing subtypes 1c, 1d, 1e, 1g, 1h, 1i, 1j and 1k, and two new subtypes 1m and 1n, and four unclassified variants reveal ancestral relationships among subtypes. The Journal of general virology. 2014;95(Pt 7):1479-87. doi: 10.1099/vir.0.064980-0

168.   Forni D, Cagliani R, Pontremoli C, Pozzoli U, Vertemara J, De Gioia L, et al. Evolutionary Analysis Provides Insight Into the Origin and Adaptation of HCV. Front Microbiol. 2018;9:854. doi: 10.3389/fmicb.2018.00854

169.   Nielsen R, Akey JM, Jakobsson M, Pritchard JK, Tishkoff S, Willerslev E. Tracing the peopling of the world through genomics. Nature. 2017;541(7637):302-10. doi: 10.1038/nature21347

170.   Li C, Cao H, Lu L, Murphy D. Full-length sequences of 11 hepatitis C virus genotype 2 isolates representing five subtypes and six unclassified lineages with unique geographical distributions and genetic variation patterns. The Journal of general virology. 2012;93(Pt 6):1173-84. doi: 10.1099/vir.0.038315-0

171.   Zhebrun AB, Kalinina OV. Viral Hepatitis C: Evolution of the epidemiologic process, evolution of the virus. Zhurnal mikrobiologii, epidemiologii i immunobiologii. 2016(1):102-12.

172.   Pybus OG, Barnes E, Taggart R, Lemey P, Markov PV, Rasachak B, et al. Genetic history of hepatitis C virus in East Asia. J Virol. 2009;83(2):1071-82. doi: 10.1128/jvi.01501-08

173.   Njouom R, Frost E, Deslandes S, Mamadou-Yaya F, Labbé AC, Pouillot R, et al. Predominance of hepatitis C virus genotype 4 infection and rapid transmission between 1935 and 1965 in the Central African Republic. The Journal of general virology. 2009;90(Pt 10):2452-6. doi: 10.1099/vir.0.011981-0

174.   Iles JC, Raghwani J, Harrison GLA, Pepin J, Djoko CF, Tamoufe U, et al. Phylogeography and epidemic history of hepatitis C virus genotype 4 in Africa. Virology. 2014;464-465(100):233-43. doi: 10.1016/j.virol.2014.07.006

175.   Forbi JC, Purdy MA, Campo DS, Vaughan G, Dimitrova ZE, Ganova-Raeva LM, et al. Epidemic history of hepatitis C virus infection in two remote communities in Nigeria, West Africa. The Journal of general virology. 2012;93(Pt 7):1410-21. doi: 10.1099/vir.0.042184-0

176.   Lyons S, Kapoor A, Sharp C, Schneider BS, Wolfe ND, Culshaw G, et al. Nonprimate hepaciviruses in domestic horses, United kingdom. Emerging infectious diseases. 2012;18(12):1976-82. doi: 10.3201/eid1812.120498

177.   Magiorkinis G, Magiorkinis E, Paraskevis D, Ho SY, Shapiro B, Pybus OG, et al. The

global spread of hepatitis C virus 1a and 1b: a phylodynamic and phylogeographic analysis. PLoS medicine. 2009;6(12):e1000198. doi: 10.1371/journal.pmed.1000198

178.   Markov PV, Pepin J, Frost E, Deslandes S, Labbe AC, Pybus OG. Phylogeography and molecular epidemiology of hepatitis C virus genotype 2 in Africa. The Journal of general virology. 2009;90(Pt 9):2086-96. doi: 10.1099/vir.0.011569-0

179.   Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, et al. Clustal W and Clustal X version 2.0. Bioinformatics. 2007;23(21):2947-8. doi: 10.1093/bioinformatics/btm404

180.   Gouy M, Guindon S, Gascuel O. SeaView version 4: A multiplatform graphical user interface for sequence alignment and phylogenetic tree building. Molecular biology and evolution. 2010;27(2):221-4. doi: 10.1093/molbev/msp259

181.   Katoh K, Rozewicki J, Yamada KD. MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. Briefings in bioinformatics. 2019;20(4):1160-6. doi: 10.1093/bib/bbx108

182.   Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Molecular biology and evolution. 2015;32(1):268-74. doi: 10.1093/molbev/msu300

183.   Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. ModelFinder: fast model selection for accurate phylogenetic estimates. Nature methods. 2017;14(6):587-9. doi: 10.1038/nmeth.4285

184.   Bozdogan H. Model selection and Akaike's Information Criterion (AIC): The general theory and its analytical extensions. Psychometrika. 1987;52(3):345-70. doi: 10.1007/BF02294361

185.   Kumar S, Stecher G, Tamura K. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. Molecular biology and evolution. 2016;33(7):1870-4. doi: 10.1093/molbev/msw054

186.   Suchard MA, Lemey P, Baele G, Ayres DL, Drummond AJ, Rambaut A. Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. Virus evolution. 2018;4(1):vey016. doi: 10.1093/ve/vey016

187.   Rambaut A, Drummond AJ, Xie D, Baele G, Suchard MA. Posterior Summarization in Bayesian Phylogenetics Using Tracer 1.7. Systematic biology. 2018;67(5):901-4. doi: 10.1093/sysbio/syy032

188.   Drummond AJ, Rambaut A, Suchard MA, Alekseyenko A, Ayres D, Baele G, et al. BEASTdoc. 2021. updated Aug 30, 2021; Available from: https://beast.community/index.html.

189.   Dellicour S, Gill MS, Faria NR, Rambaut A, Pybus OG, Suchard MA, et al. Relax, Keep Walking - A Practical Guide to Continuous Phylogeographic Inference with BEAST. Molecular biology and evolution. 2021;38(8):3486-93. doi: 10.1093/molbev/msab031

190.   Kalaghatgi P, Sikorski AM, Knops E, Rupp D, Sierra S, Heger E, et al.

Geno2pheno[HCV] - A Web-based Interpretation System to Support Hepatitis C Treatment Decisions in the Era of Direct-Acting Antiviral Agents. PloS one. 2016;11(5):e0155869. doi: 10.1371/journal.pone.0155869

191.   Silvestro D, Michalak I. raxmlGUI: a graphical front-end for RAxML. Organisms Diversity & Evolution. 2012;12(4):335-7. doi: 10.1007/s13127-011-0056-0

192.   Shapiro B, Rambaut A, Drummond AJ. Choosing appropriate substitution models for the phylogenetic analysis of protein-coding sequences. Molecular biology and evolution. 2006;23(1):7-9. doi: 10.1093/molbev/msj021

193.   Gray RR, Parker J, Lemey P, Salemi M, Katzourakis A, Pybus OG. The mode and tempo of hepatitis C virus evolution within and among hosts. BMC evolutionary biology. 2011;11:131. doi: 10.1186/1471-2148-11-131

194.   Al-Qahtani AA, Baele G, Khalaf N, Suchard MA, Al-Anazi MR, Abdo AA, et al. The epidemic dynamics of hepatitis C virus subtypes 4a and 4d in Saudi Arabia. Scientific reports. 2017;7:44947. doi: 10.1038/srep44947

195.   Brandão R, Marcelino R, Gonçalves F, Diogo I, Carvalho A, Cabanas J, et al. Characterization of NS5A and NS5B Resistance-Associated Substitutions from Genotype 1 Hepatitis C Virus Infected Patients in a Portuguese Cohort. Viruses. 2018;10(5):223. doi: 10.3390/v10050223

196.   Rambaut A, Lam TT, Max Carvalho L, Pybus OG. Exploring the temporal structure of heterochronous sequences using TempEst (formerly Path-O-Gen). Virus evolution. 2016;2(1):vew007. doi: 10.1093/ve/vew007

197.   Palladino C, Ezeonwumelu IJ, Marcelino R, Briz V, Moranguinho I, Serejo F, et al. Epidemic history of hepatitis C virus genotypes and subtypes in Portugal. Scientific reports. 2018;8(1):12266. doi: 10.1038/s41598-018-30528-0

198.   Guirgis BS, Abbas RO, Azzazy HM. Hepatitis B virus genotyping: current methods and clinical implications. International journal of infectious diseases : IJID : official publication of the International Society for Infectious Diseases. 2010;14(11):e941-53. doi: 10.1016/j.ijid.2010.03.020

199.   SiemensDiagnostics. TRUGENE HBV Genotyping Assay. Published by: Siemens Healthineers group; 2009. Available from: https://www.labmedica.com/whitepapers/Trugene_HBV_Assay_Final_cover_8.28.pdf

200.   Office_of_Studies_and_Strategy. Resident Foreign Population in Portugal - TOTAL. Portuguese Republic - Economy and Digital Transition; 2021. Available from: https://www.gee.gov.pt/pt/lista-publicacoes/estatisticas-de-imigrantes-em-portugal-por-nacionalidade/regioes-do-mundo/GEE%20-%20Popula%C3%A7%C3%A3o%20estrangeira%20residente%20em%20Portugal%20-%20Mundo.pdf/9129-populacao-estrangeira-residente-em-portugal-mundo/file.

201.   Office_of_Studies_and_Strategy. Resident Foreign Population in Portugal - Europe. Portuguese Republic - Economy and Digital Transition; 2021. Available from:

https://www.gee.gov.pt/pt/lista-publicacoes/estatisticas-de-imigrantes-em-portugal-por-nacionalidade/regioes-do-mundo/GEE%20-%20Popula%C3%A7%C3%A3o%20estrangeira%20residente%20em%20Portugal%20-%20Europa.pdf/9127-populacao-estrangeira-com-estatuto-legal-de-residente-em-portugal-europa/file.

202.   Office_of_Studies_and_Strategy. Resident Foreign Population in Portugal - Africa. Portuguese Republic - Economy and Digital Transition; 2021. Available from: https://www.gee.gov.pt/pt/lista-publicacoes/estatisticas-de-imigrantes-em-portugal-por-nacionalidade/regioes-do-mundo/GEE%20-%20Popula%C3%A7%C3%A3o%20estrangeira%20residente%20em%20Portugal%20-%20%C3%81frica.pdf/9126-populacao-estrangeira-com-estatuto-legal-de-residente-em-portugal-africa/file.

203.   Office_of_Studies_and_Strategy. Resident Foreign Population in Portugal - America. Portuguese Republic - Economy and Digital Transition; 2021. Available from: https://www.gee.gov.pt/pt/lista-publicacoes/estatisticas-de-imigrantes-em-portugal-por-nacionalidade/regioes-do-mundo/GEE%20-%20Popula%C3%A7%C3%A3o%20estrangeira%20residente%20em%20Portugal%20-%20Am%C3%A9rica.pdf/9125-populacao-estrangeira-com-estatuto-legal-de-residente-em-portugal-america/file.

204.   Góis P, Marques JC. Retrato de um Portugal migrante: a evolução da emigração, da imigração e do seu estudo nos últimos 40 anos. 2018. Vol. 29. doi: 10.4000/eces.3307. Available from: http://journals.openedition.org/eces/3307.

205.   Office_of_Studies_and_Strategy. Resident Foreign Population in Portugal - Ukraine. Portuguese Republic - Economy and Digital Transition; 2021. Available from: https://www.gee.gov.pt/pt/lista-publicacoes/estatisticas-de-imigrantes-em-portugal-por-nacionalidade/paises/Ucr%C3%A2nia/3992-populacao-estrangeira-com-estatuto-legal-de-residente-em-portugal-ucrania/file.

206.   Office_of_Studies_and_Strategy. Resident Foreign Population in Portugal - Brazil. Portuguese Republic - Economy and Digital Transition; 2021. Available from: https://www.gee.gov.pt/pt/lista-publicacoes/estatisticas-de-imigrantes-em-portugal-por-nacionalidade/paises/Brasil/4017-populacao-estrangeira-com-estatuto-legal-de-residente-em-portugal-brasil/file.

207.   Office_of_Studies_and_Strategy. Resident Foreign Population in Portugal - Cape Verde. Portuguese Republic - Economy and Digital Transition; 2021. Available from: https://www.gee.gov.pt/pt/lista-publicacoes/estatisticas-de-imigrantes-em-portugal-por-nacionalidade/paises/Cabo%20Verde/3947-populacao-estrangeira-com-estatuto-legal-de-residente-em-portugal-cabo-verde/file.

208.   de Pina-Araujo IIM, Spitz N, Soares CC, Niel C, Lago BV, Gomes SA. Hepatitis B virus genotypes A1, A2 and E in Cape Verde: Unequal distribution through the islands and association with human flows. PloS one. 2018;13(2):e0192595. doi: 10.1371/journal.pone.0192595

References

209.   Ciccozzi M, Babakir-Mina M, Lo Presti A, Salpini R, Cella E, Gabanelli E, et al. Molecular analysis of hepatitis B virus in Bulgaria. Journal of medical virology. 2013;85(1):49-54. doi: 10.1002/jmv.23432

210.   Lo Presti A, Andriamandimby SF, Lai A, Angeletti S, Cella E, Mottini G, et al. Origin and evolutionary dynamics of Hepatitis B virus (HBV) genotype E in Madagascar. Pathogens and global health. 2017;111(1):23-30. doi: 10.1080/20477724.2016.1278103

211.   Zhou Y, Holmes EC. Bayesian estimates of the evolutionary rate and age of hepatitis B virus. Journal of molecular evolution. 2007;65(2):197-205. doi: 10.1007/s00239-007-0054-1

212.   Hespanhol L, Vallio CS, Costa LM, Saragiotto BT. Understanding and interpreting confidence and credible intervals around effect estimates. Brazilian journal of physical therapy. 2019;23(4):290-301. doi: 10.1016/j.bjpt.2018.12.006

213.   Kostaki EG, Karamitros T, Stefanou G, Mamais I, Angelis K, Hatzakis A, et al. Unravelling the history of hepatitis B virus genotypes A and D infection using a full-genome phylogenetic and phylogeographic approach. eLife. 2018;7::e36709. doi: 10.7554/eLife.36709

214.   Kalkauskas A, Perron U, Sun Y, Goldman N, Baele G, Guindon S, et al. Sampling bias and model choice in continuous phylogeography: Getting lost on a random walk. PLoS computational biology. 2021;17(1):e1008561. doi: 10.1371/journal.pcbi.1008561

215.   Pereira MH. A emigração portuguesa para o Brasil e a geo-estratégia do desenvolvimento euro-americano. 2009. Vol. 56. doi: 10.4000/lerhistoria.1925. Available from: http://journals.openedition.org/lerhistoria/1925.

216.   Gisselquist DP. Estimating HIV-1 transmission efficiency through unsafe medical injections. International journal of STD & AIDS. 2002;13(3):152-9. doi: 10.1258/0956462021924820

217.   Harrison A, Lemey P, Hurles M, Moyes C, Horn S, Pryor J, et al. Genomic analysis of hepatitis B virus reveals antigen state and genotype as sources of evolutionary rate variation. Viruses. 2011;3(2):83-101. doi: 10.3390/v3020083

218.   Bouckaert R, Alvarado-Mora MV, Pinho JR. Evolutionary rates and HBV: issues of rate estimation with Bayesian molecular methods. Antiviral therapy. 2013;18(3 Pt B):497-503. doi: 10.3851/imp2656

219.   Navascués M, Leblois R, Burgarella C. Demographic inference through approximate-Bayesian-computation skyline plots. PeerJ. 2017;5:e3530. doi: 10.7717/peerj.3530

220.   Kartashev V, Doring M, Nieto L, Coletta E, Kaiser R, Sierra S, et al. New findings in HCV genotype distribution in selected West European, Russian and Israeli regions. J Clin Virol. 2016;81:82-9. doi: 10.1016/j.jcv.2016.05.010

221.   Petruzziello A, Marigliano S, Loquercio G, Cozzolino A, Cacciapuoti C. Global epidemiology of hepatitis C virus infection: An up-date of the distribution and circulation of hepatitis C virus genotypes. World journal of gastroenterology. 2016;22(34):7824-40. doi:

10.3748/wjg.v22.i34.7824

222.   Choo QL, Kuo G, Weiner AJ, Overby LR, Bradley DW, Houghton M. Isolation of a cDNA clone derived from a blood-borne non-A, non-B viral hepatitis genome. Science. 1989;244(4902):359-62. doi: 10.1126/science.2523562

223.   Kuo G, Choo QL, Alter HJ, Gitnick GL, Redeker AG, Purcell RH, et al. An assay for circulating antibodies to a major etiologic virus of human non-A, non-B hepatitis. Science. 1989;244(4902):362-4. doi: 10.1126/science.2496467

224.   Moreira M, Hughes B, Costa Storti C, Zobel F. Drug Policy Profile: Portugal. Published by: EMCDDA; 2011.

225.   Trends in injecting drug use in Europe. Published by: EMCDDA; 2010. Available from: https://www.emcdda.europa.eu/publications/selected-issues/injecting_en

226.   Csete J, Kamarulzaman A, Kazatchkine M, Altice F, Balicki M, Buxton J, et al. Public health and international drug policy. Lancet (London, England). 2016;387(10026):1427-80. doi: 10.1016/S0140-6736(16)00619-X

227.   Koletzki D, Dumont S, Vermeiren H, Peixe P, Nina J, Camacho RJ, et al. Full genome sequence of three isolates of hepatitis C virus subtype 4b from Portugal. Archives of virology. 2009;154(1):127-32. doi: 10.1007/s00705-008-0270-z

228.   Wiessing L, Ferri M, Grady B, Kantzanou M, Sperle I, Cullen KJ, et al. Hepatitis C virus infection epidemiology among people who inject drugs in Europe: a systematic review of data for scaling up treatment and prevention. PloS one. 2014;9(7):e103345. doi: 10.1371/journal.pone.0103345

229.   Feucht HH, Schroter M, Zollner B, Polywka S, Nolte H, Laufs R. The influence of age on the prevalence of hepatitis C virus subtypes 1a and 1b. The Journal of infectious diseases. 1997;175(3):685-8. doi: 10.1093/infdis/175.3.685

230.   Bracho MA, Saludes V, Martro E, Bargallo A, Gonzalez-Candelas F, Ausina V. Complete genome of a European hepatitis C virus subtype 1g isolate: phylogenetic and genetic analyses. Virol J. 2008;5:72. doi: 10.1186/1743-422X-5-72

231.   De Luca A, Di Giambenedetto S, Lo Presti A, Sierra S, Prosperi M, Cella E, et al. Two Distinct Hepatitis C Virus Genotype 1a Clades Have Different Geographical Distribution and Association With Natural Resistance to NS3 Protease Inhibitors. Open forum infectious diseases. 2015;2(2):ofv043. doi: 10.1093/ofid/ofv043

232.   Santos AF, Bello G, Vidal LL, Souza SL, Mir D, Soares MA. In-depth phylogenetic analysis of hepatitis C virus subtype 1a and occurrence of 80K and associated polymorphisms in the NS3 protease. Scientific reports. 2016;6:31780. doi: 10.1038/srep31780

233.   Palladino C, Sanchez-Carrillo M, Mate-Cano I, Vazquez-Moron S, Jimenez-Sousa MA, Gutierrez-Rivas M, et al. Low frequency of NS5A relevant resistance-associated substitutions to Elbasvir among hepatitis C virus genotype 1a in Spain: a cross-sectional study. Sci Rep. 2017;7(1):2892. doi: 10.1038/s41598-017-02968-7

234.   Palladino C, Ezeonwumelu IJ, Mate-Cano I, Borrego P, Martínez-Román P, Arca-

Lafuente S, et al. Epidemic history and baseline resistance to NS5A-specific direct acting drugs of hepatitis C virus in Spain. Scientific reports. 2020;10(1):13024. doi: 10.1038/s41598-020-69692-7

235. Simmonds P. Variability of hepatitis C virus. Hepatology (Baltimore, Md). 1995;21(2):570-83. doi: 10.1002/hep.1840210243

236. Mellor J, Holmes EC, Jarvis LM, Yap PL, Simmonds P. Investigation of the pattern of hepatitis C virus sequence diversity in different geographical regions: implications for virus classification. The International HCV Collaborative Study Group. The Journal of general virology. 1995;76 (Pt 10):2493-507. doi: 10.1099/0022-1317-76-10-2493

237. Baiôa M, Fernandes PJ, de Meneses FR. The Political History of Twentieth-Century Portugal. 2003. Vol. 1 (2). Available from: https://www.cepese.pt/portal/pt/publicacoes/obras/volume-1-number-2-winter-2003/the-political-history-of-twentieth-century-portugal.

238. Bover O, Velilla P. Migration in Spain: Historical background and current trends. In: IZA discussion paper n.º 88. CEPR conference on European migration; Munich, Germany. 1999.

239. Domosławski A. Drug Policy in Portugal: The Benefits of Decriminalizing Drug Use. Published by: Open Society Foundations; 2011. Available from: https://www.opensocietyfoundations.org/publications/drug-policy-portugal-benefits-decriminalizing-drug-use

240. Melo e Costa C. Breve História da Transfusão: transfusão de concentrado eritrocitário no doente crítico (Translation: Brief History of Transfusion: transfusion of erythrocyte concentrate in the critically ill). Published in: Repository of Hospital Dr. Prof. Fernando Fonseca; 2010. Available from: https://repositorio.hff.min-saude.pt/bitstream/10400.10/1278/1/Carolina%20Costa-%20Historia%20da%20transfusao.pdf

241. Lusa-Agency. Hepatite, a última emboscada no Ultramar (Translation: Hepatitis, the last ambush overseas). País; 2006. Available from: https://www.rtp.pt/noticias/pais/hepatite-a-ultima-emboscada-no-ultramar_n35628.

242. Tadeu TAA. A Guarda durante a II Guerra Mundial (Translation: The Guarda during World War II). [MsC Thesis]. Published in: Repository of University of Coimbra, Portugal; 2011. Available from: http://hdl.handle.net/10316/18164

243. Lopes MMD. MEMÓRIAS DO SALTO: memórias da emigração ilegal para França, entre 1954 e 1974 (Translation: MEMORIES OF THE JUMP: memories of illegal emigration to France, between 1954 and 1974). [MsC Thesis]. Published in Faculdade de Letras da Universidade do Porto; 2016. Available from: https://repositorio-aberto.up.pt/handle/10216/87380

244. Baganha MIB. Portuguese Emigration After World War II. Published by Wordpress; 2015. p: 1-19. Available from: https://emigratecaportuguesa.files.wordpress.com/2015/04/portuguese-emigration-after-

world-war-ii.pdf.

245.   Baganha MIB, Góis P, Pereira PT. European Migration: What do we know? Chapter 12. Published by Oxford University Press; 2005. p. 415-58. Available from: http://www3.uma.pt/ppereira/Eumi-121.pdf.

246.   Gonçalves C. O século dos italianos na Suíça (Translation: The century of Italians in Switzerland); 2003. Available from: https://www.swissinfo.ch/por/o-s%C3%A9culo-dos-italianos-na-su%C3%AD%C3%A7a/879420.

247.   Pereira V. A Ditadura de Salazar e a Emigração - O Estado português e os seus emigrantes em França (1957-1974) [Translation: The Salazar Dictatorship and Emigration - The Portuguese State and its emigrants in France (1957-1974)]. Published by Temas e Debates; 2014.

248.   EASL. EASL recommendations on treatment of hepatitis C: Final update of the series(☆). Journal of hepatology. 2020;73(5):1170-218. doi: 10.1016/j.jhep.2020.08.018

249.   Childs K, Davis C, Cannon M, Montague S, Filipe A, Tong L, et al. Suboptimal SVR rates in African patients with atypical genotype 1 subtypes: Implications for global elimination of hepatitis C. Journal of hepatology. 2019;71(6):1099-105. doi: 10.1016/j.jhep.2019.07.025

250.   Pawlotsky JM. Hepatitis C Virus Resistance to Direct-Acting Antiviral Drugs in Interferon-Free Regimens. Gastroenterology. 2016;151(1):70-86. doi: 10.1053/j.gastro.2016.04.003

251.   Di Lello FA, Culasso AC, Campos RH. Inter and intrapatient evolution of hepatitis C virus. Annals of hepatology. 2015;14(4):442-9. doi: 10.1016/S1665-2681(19)31164-0

252.   Weaver SC, Forrester NL, Liu J, Vasilakis N. Population bottlenecks and founder effects: implications for mosquito-borne arboviral emergence. Nature reviews Microbiology. 2021;19(3):184-95. doi: 10.1038/s41579-020-00482-8

253.   Smith D, Magri A, Bonsall D, Ip CLC, Trebes A, Brown A, et al. Resistance analysis of genotype 3 hepatitis C virus indicates subtypes inherently resistant to nonstructural protein 5A inhibitors. Hepatology (Baltimore, Md). 2019;69(5):1861-72. doi: 10.1002/hep.29837

254.   Fridell RA, Qiu D, Wang C, Valera L, Gao M. Resistance analysis of the hepatitis C virus NS5A inhibitor BMS-790052 in an in vitro replicon system. Antimicrob Agents Chemother. 2010;54(9):3641-50. doi: 10.1128/aac.00556-10

255.   Lontok E, Harrington P, Howe A, Kieffer T, Lennerstrand J, Lenz O, et al. Hepatitis C virus drug resistance-associated substitutions: State of the art summary. Hepatology (Baltimore, Md). 2015;62(5):1623-32. doi: 10.1002/hep.27934

256.   Hayes CN, Imamura M, Chayama K. Management of HCV patients in cases of direct-acting antiviral failure. Expert review of gastroenterology & hepatology. 2019;13(9):839-48. doi: 10.1080/17474124.2019.1651642

257.   P. Krishnan, R. Tripathi, G. Schnell, T. Reisch, J. Beyer, T. Dekhtyar, et al., editors. Long-term follow-up of treatment emegent resistance associated variants in NS3, NS5A and

NS5B with Paritaprevir, Ombitasvir and Dasabuvir based regimens (Oral presentation). Conference on Retroviruses and Opportunistic Infections; 2015.

258.   Sulkowski M, Krishnan P, Tripathi R, Schnell G, Xie Y, Cohen D, et al. Effect of Baseline Resistance-Associated Variants on SVR With the 3D Regimen Plus RBV (Poster presentation). Conference on Retroviruses and Opportunistic Infections; 2016.

259.   Buti M, Gordon SC, Zuckerman E, Lawitz E, Calleja JL, Hofer H, et al. Grazoprevir, Elbasvir, and Ribavirin for Chronic Hepatitis C Virus Genotype 1 Infection After Failure of Pegylated Interferon and Ribavirin With an Earlier-Generation Protease Inhibitor: Final 24-Week Results From C-SALVAGE. Clinical infectious diseases : an official publication of the Infectious Diseases Society of America. 2016;62(1):32-6. doi: 10.1093/cid/civ722

260.   Zeuzem S, Ghalib R, Reddy KR, Pockros PJ, Ben Ari Z, Zhao Y, et al. Grazoprevir-Elbasvir Combination Therapy for Treatment-Naive Cirrhotic and Noncirrhotic Patients With Chronic Hepatitis C Virus Genotype 1, 4, or 6 Infection: A Randomized Trial. Ann Intern Med. 2015;163(1):1-13. doi: 10.7326/m15-0785

SUPPLEMENTS

ANNEX 1

Details about HBV DNA extraction, amplification, and sequencing of pol gene

HBV DNA extraction (QIAamp DNA Blood Mini Kit, QIAGen, Werfen)

a) Reagent's preparation when opening the kit for the first time

- Protease: Add 5.5 ml of protease solvent included in the kit (nuclease-free water containing 0.04% sodium azide) to the tube containing the lyophilized protease. Reconstituted protease is stable for up to two months when stored at 2-8°C. Storage at -20ºC is recommended to prolong the life of the protease. To avoid successive freezing and thawing, store protease in aliquots.

- Buffer AW1: add 125 ml of 96-100% ethanol to 95 ml of concentrated Buffer AW1 and mix by inversion 5-10 times.

- Buffer AW2: pipette 160 ml of 96-100% ethanol into 66 ml of buffer AW2 and mix by inversion 5-10 times.
  Note: Unused volume of reconstituted Buffer AW1 and Buffer AW2 can be stored at room temperature for 1 year.

b) HBV DNA extraction

- Heat a thermal block to 56ºC.

- Shake AL buffer before using it and if there is any precipitate, dissolve it in a bath at 56ºC.

- Prepare 350 µl of negative extraction control from a pool of negative plasmas for this virus.

- Pipette 350 µl of HBV plasma sample and positive and negative controls into 1.5ml microtubes.

- Add 35 µl of the reconstituted protease into the microtubes.

- Homogenize the AL buffer and add 350 µl to that same microtubes.

- Vortex intermittently for 15 seconds; to promote an efficient lysis it is essential that the sample and the AL buffer are mixed until a homogeneous solution is obtained.

- Add 3.5 µl carrier DNA to the cap of each tube, close the tubes and mix by inversion.

  Note: carrier DNA enables viral DNA binding to the column membrane on the next step. It is not provided in the kit. Here it was used Polyadenylic Acid, PolyA, 10 mg/ml; Amersham; It should be stored at -20°C.

- Vortex intermittently for 15 seconds and spin the microtubes in the centrifuge for approximately 4 seconds.

- Incubate the tubes in the thermal block at 56°C for 10 minutes; the amount of DNA released is maximum after lysis during this incubation step. Longer incubation periods have no effect on the extraction or quality of purified DNA.

- Spin the microtubes in the centrifuge for approximately 4 seconds.

- Add 402.5µl 96-100% ethanol, vortex intermittently for 15 seconds and then spin in centrifuge.

- Apply the mixture from the previous step (630µl at a time), carefully, in a QIAamp Mini column incorporated into a 2ml collection tube, without letting the mixture reach the edge of the column, close the lid and centrifuge at 6000g for 1 minute.

- Place the column in a 2ml new collection tube and discard the tube containing the filtrate.

- Repeat the two previous steps.

- Add 500µl of buffer AW1, close the lid and centrifuge at 6000g for 1 minute.

- Place the column in a new 2ml collection tube and discard the tube containing the filtrate.

- Add 500µl of buffer AW2, close the column cover and centrifuge at 20000g for 3 minutes.

- Place the column in a 2ml microtube (not supplied in the kit) and discard the tube containing the filtrate.

- Centrifuge at 20000g for 1 minute to completely dry the membrane.

- Discard the 2 ml tube containing the filtrate and introduce the column into a previously identified 2 ml screw tube for elution of virus DNA.

- Apply 60µl of buffer AE in the center of the membrane, close the column lid and incubate at room temperature for 5 minutes.

- Centrifuge at 6000g for 1 minute.

- Samples can be stored for 24 hours at 2-8°C, at -20°C for longer periods or can be used immediately.

<u>Amplification of HBV Pol gene</u> (Biotaq DNA polimerase, 5 U/µl; Bioline/Meridian BioScience)

a) 1st round PCR

- Put the reagents of amplification kit at room temperature, shake them slightly, spin them and place them on ice.
  Note: the enzyme should only be removed from the freezer on the moment it is going to be used.

- Mark 0.2 ml microtubes for each of the samples, for the positive control, for the negative control of extraction and for the negative control of PCR.

- Prepare the mixture for the PCR reaction in a suitable microtube.

PCR mixture (for 1 sample)

Water nuclease free - 14.45 µl

NH4 buffer (10x) - 2.5 µl

dNTPs mix (100mM) - 0.2 µl

FW_O (10mM) - 1 µl

RV_O (10mM) - 1 µl

$MgCl_2$ (50mM) - 0.75 µl

BioTaq enzyme - 0.1 µl

- Distribute 20 µl of the mixture into each of the 0.2 ml microtubes (include one extra tube for negative control of PCR technique).

- Add 5 µl of each of the extracted samples and controls to the respective tubes. Add 5 µl of "nuclease free" water int the PCR negative control.

- Place the tubes in a thermal cycler under the following conditions:

167

94ºC	4 min	
94ºC	45 secs	
55ºC	30 secs	40 cycles
72ºC	1 min (+5 secs/cycle)	
72ºC	15 min	
4ºC	∞	

- After the end of the PCR reaction, save the amplified products at -20ºC or proceed with the nested PCR

b) 2nd round PCR

- Proceed exactly as for the first PCR, for reagents and microtubes, also using the same amplification kit.
- Prepare the mixture for the PCR reaction in a 1.5 ml microtube.

PCR mixture (for 1 sample)

Water nuclease free - 18.45 µl

NH4 buffer (10x) - 2.5 µl

dNTPs mix (100mM) - 0.2 µl

FW_I (10mM) - 1 µl

RV_I (10mM) - 1 µl

MgCl2 (50mM) - 0.75 µl

BioTaq enzyme - 0.1 µl

- Distribute 24 µl of the mixture into each of the 0.2 ml microtubes.

- Add 1µl of each of the 1st round PCR products to the respective tubes.

Note: The product obtained for the positive control in the first PCR must be diluted 1:5 (4µl H_2O + 1µl product).

- Put the 0.2ml microtubes in a thermocycler under the same cycling conditions used for the 1st round PCR.

Agarose gel electrophoresis

- Prepare a 1.5% agarose gel.
- Fill an electrophoresis cuvette with 1x TAE buffer.
- Prepare the samples to be applied to the gel, adding 10µl of each to 2µl of Orange G dye.
- To 0.5µl of the molecular weight marker (GeneRuler 100 bp DNA Ladder, 0,5 µg/µl; ThermoFisher Scientific), 2µl of the dye and 7.5µl of water are added to make a volume of 10µl.
- Apply the stained samples into the wells of the agarose gel and apply electric current (70 Volts) for approximately 30 minutes.
- Observe the gel in a transilluminator where a 943 bp band should be visualized for samples and positive control.

PCR products purification using a spin column system (JetQuick PCR Product Purification Spin Kit; Genomed)

- Place a 1.5 ml microtube with nuclease free water to heat at 65ºC.
- Add 60µl of H1 kit solution to remaining 15µl of 2nd round PCR product.
- Mix and transfer the mixture to a purification column previously inserted in a supernatant collection tube.
- Centrifuge at 15000g for 1 minute.
- Discard the contents of the collection tube.
- Add 500µl of H2 solution to the column.
- Repeat the above centrifugation and discard steps.

- Centrifuge at 16000g for 1 minute to completely dry the column membrane.

- Insert the column into a 1.5ml microtube.

- Add 20-35µl of nuclease free water at 65ºC to each column.

  Note: Add a higher volume, up to 50µl, if the band obtained after electrophoresis is too strong.

- Wait 1 minute.

- Centrifuge at 16000g for 2 minutes.

- Store purified products at -20°C or proceed with the sequencing reaction.

Sequencing Reaction (BigDye terminator v.3.1 Cycle Sequencing Kit; Applied Biosystems)

- Thaw the reagents needed for the sequencing reactions and the purified PCR products:

  BigDye terminator v.3.1 (Applied Biosystems)

  BigDye terminator 5X sequencing buffer (Applied Biosystems)

  Primers Forward and Reverse from the nested PCR (5µM)

  Nuclease free water

- Place the thawed reagents on ice and prepare a mixture to each primers as described below:

Sequencing mixture (for 1 sample)

BigDye terminator v.3.1 - 0.5 µl

Primer (5µM) - 1 µl

Buffer (5X) - 1.75 µl

- Mix the reagents by pipetting up and down several times and make a spin to the microtube.

- Distribute 3.25 µl of each mix into the two different tubes.

- Add between 0.3 - 1 µl of the previously purified PCR product to each tube, according to the intensity of the band observed on the agarose gel.
- Complete the reaction volume to 10 µl with nuclease free water and place the samples in the thermocycler under the following conditions:

Sequencing cycle

96ºC	1 min	
96ºC	10 secs	25 cycles
60ºC	4 min	
4ºC	7 min	
4ºC	∞	

Sequencing products purification through Ethanol/Sodium Acetate

- Prepare a sodium acetate/ethanol solution by mixing 2 µl of 3M sodium acetate (pH 4.6) with 50 µl of 100% ethanol, per reaction.
- Add 52 µl of the previous solution to each sequencing reaction and transfer to a 96 well plate.
- Cover the plate with aluminium foil.
- Vortex the plate to mix.
- Centrifuge the plate at 2000g for 20 minutes.
- Once the centrifugation is finished, remove the aluminium foil carefully trying to not disturb the sediment.
- Immediately place several absorbent paper towels without residues on top of the plate and invert it.
- Centrifuge the inverted plate at 150g for 1 minute.
- Add 150 µl of 70% ethanol, prepared previously, and cover the plate with a new aluminium foil.
- Centrifuge at 2000g for 5 minutes.

- When the centrifugation is finished, remove again the aluminium foil carefully trying to not disturb the sediment.
- Invert the plate over clean towels and centrifuge the inverted plate at 150g for 1 minute.
- After centrifuging, allow the plate to air dry for approximately 1 minute to eliminate any remaining ethanol.
- Cover the plate with aluminium foil to store between –15ºC and –25ºC in the dark, but no longer than 2 weeks or add 20 µl of Hi-Di^{TM} Formamide (Applied Biosystems, USA) and move to the processing phase of the sequences in the ABI Prism 3100-Avant Genetic Analyzer according with the equipment manual.

ANNEX 2

Table A2. HBV reference sequence details used for the spatio-temporal analysis: origin country, collection date and accession number.

Genotype/Subgenotype	Country	Collection Date	Accession Number
A	Cameroon	1994	AB194952
A	Ghana	2000	GQ161753
A	Gabon	2004	FJ349296
A	Nigeria	2004	FJ692554
A	Nigeria	2004	KY494150
A	Gabon	2005	EU054331
A	Haiti	2006	FJ692593
A	Nigeria	2007	HM363612
A	Panama	2010	KJ638668
A	Cameroon	2011	KY493906
A	France	2011	KP712829
A	Cameroon	2012	KY493834
A	Cameroon	2012	KY493855
A	Cameroon	2013	KY493861
A	Cuba	2013	KM606737
A	Ghana	2013	KU711605
A	Haiti	2013	KP234053
A	Cameroon	2014	KY493820
A	Cameroon	2014	KY493872
A	Cameroon	2015	KY494162
A	Kenya	2018	MK127853
A	Cameroon	1998_2004	AM180624
A	Cameroon	2006_2007	FN545836
A	USA	2006_2009	JN604217
A1	Argentina	2010	KJ810927
A1	Bangladesh	2007	JQ514505
A1	Bangladesh	2015	MF925372
A1	Belgium	1998	GQ331047
A1	Belgium	2000	GU563548
A1	Belgium	2003	GU563547
A1	Botswana	2009	MH464828
A1	Botswana	2014	KX648548
A1	Botswana	2015	MF979146
A1	Brazil	1965	KX264620
A1	Brazil	1969	KX264597
A1	Brazil	1970	KX264656
A1	Brazil	1974	KX264607
A1	Brazil	1974	KX264619
A1	Brazil	1976	KX264598
A1	Brazil	1981	KX264658
A1	Brazil	1983	KX264612
A1	Brazil	1984	KX264638
A1	Brazil	1984	KX264635
A1	Brazil	1987	KX264595
A1	Brazil	1988	KX264659
A1	Brazil	2001	KU847526
A1	Brazil	2005	KJ854685
A1	Brazil	2007	KJ854694
A1	Brazil	2008	KY809927
A1	Brazil	2008	KY809924
A1	Brazil	2009	JN983892
A1	Brazil	2010	JN983834
A1	Brazil	2011	KU847482
A1	Brazil	2012	KU847743
A1	Brazil	2013	KU847725
A1	Brazil	2016	MK517517
A1	CapeVerde	2011	MF772353
A1	CapeVerde	2016	MF772351
A1	Cuba	2006	KM606748
A1	Cuba	2013	KM606706
A1	France	2006	GQ486317
A1	France	2011	KU847537
A1	France	2011	KP712845
A1	Germany	2006	GQ486686
A1	Haiti	2006	FJ692563
A1	India	1963	KF214660
A1	India	1963	KF214656
A1	India	2014	KM590923
A1	India	2009_2012	KC875255
A1	India	2011_2012	KJ958467
A1	Italy	2008	KP997702
A1	Italy	2009	KX533510
A1	Italy	2010	KT862878
A1	Japan	1989	AB937795
A1	Japan	1999	AB937796
A1	Japan	2005_2006	AB453986
A1	Japan	2005_2006	AB453989
A1	Kenya	2012	KP168422
A1	Kenya	2018	MK127854
A1	Kenya	2018	MK127848
A1	Malawi	1997	AB076678
A1	Mozambique	2015	MF615984
A1	Netherlands	2005	GQ486535
A1	Panama	2011	KJ638661
A1	Panama	2012	KP718069
A1	Phillipines	1990_1995	M57663
A1	RepCongo	2013_2014	MH253772
A1	Rwanda	2014	MK512465
A1	Rwanda	2001_2006	FM199978
A1	SouthAfrica	2005	KF476017
A1	SouthAfrica	2006	KM374985
A1	SouthAfrica	2007	KF922428
A1	SouthAfrica	2007	KT347091
A1	SouthAfrica	2009	KJ010778
A1	SouthAfrica	2009	JN182319
A1	SouthAfrica	2010	JN182329
A1	SouthAfrica	2011	KT347092
A1	Sudan	2009	KU736920
A1	Uruguay	2013	KJ586810
A1	USA	2005	GQ486143
A1	USA	2006	GQ486271
A1	USA	2007	GQ486828
A1	USA	2008	JN604145
A1	Zimbabue	2014	KX648549

A2	Japan	1988	AB937798
A2	Russia	1993	JX125368
A2	Belgium	1995	KT749848
A2	Belgium	1995	KT749829
A2	Belgium	1995	KT749848
A2	Japan	1996	KC836881
A2	Belgium	1997	KT749824
A2	SouthAfrica	1997	JX507080
A2	Belgium	1998	EU859955
A2	USA	1998	KF779252
A2	USA	1998	KF779210
A2	Belgium	1999	KT749831
A2	Belgium	1999	EU859950
A2	USA	1999	KF779259
A2	USA	1999	KF779211
A2	Belgium	2000	KT749835
A2	Belgium	2000	EU859945
A2	USA	2000	JQ707636
A2	USA	2000	KF779268
A2	USA	2000	KF779264
A2	Argentina	2001	KJ843214
A2	Belgium	2001	GU563553
A2	USA	2001	KF779386
A2	USA	2001	KF779221
A2	Belgium	2002	KT749842
A2	Japan	2002	AB697496
A2	Argentina	2003	KJ843216
A2	Japan	2003	AB697487
A2	USA	2003	KF779231
A2	Argentina	2004	KJ843215
A2	Belgium	2004	GU563550
A2	Italy	2004	KP997576
A2	USA	2004	KF779238
A2	Belarus	2005	EU414133
A2	Italy	2005	KP997971
A2	Latvia	2005	JX096952
A2	SouthAfrica	2005	KM391914
A2	SouthAfrica	2005	KF475997
A2	USA	2005	GQ486642
A2	Belgium	2006	KT749836
A2	Belgium	2006	KT749838
A2	Cuba	2006	KM606746
A2	CzechRepub	2006	GQ486360
A2	CzechRepub	2006	GQ486232
A2	France	2006	GQ486063
A2	Germany	2006	GQ486668
A2	Poland	2006	GQ477466
A2	Poland	2006	GQ477496
A2	Poland	2006	GQ486567
A2	Poland	2006	GQ486616
A2	Poland	2006	GQ477489
A2	Poland	2006	GQ477469
A2	Poland	2006	GQ477491
A2	Poland	2006	GQ477464
A2	Tajikistan	2006	AB330372
A2	USA	2006	GQ486725
A2	USA	2006	GQ486603
A2	USA	2006	JN604269
A2	USA	2006	JN604172
A2	Belgium	2007	KT749849
A2	Brazil	2007	KY809884
A2	Brazil	2007	KY809923
A2	Brazil	2007	KY809925
A2	France	2007	GQ486691
A2	Germany	2007	GQ486710
A2	Spain	2007	GQ486829
A2	USA	2007	GQ486754
A2	USA	2007	GQ486751
A2	USA	2007	JN604266
A2	USA	2007	GQ486807
A2	Brazil	2008	KY809902
A2	USA	2008	JN604257
A2	USA	2008	JN604304
A2	Belgium	2009	KT749840
A2	Italy	2009	KP997973
A2	Italy	2009	KT288336
A2	Italy	2009	KT275271
A2	Italy	2009	KP997772
A2	Japan	2009	AB549213
A2	Italy	2010	KP997741
A2	Italy	2010	KT275266
A2	Italy	2010	KP997828
A2	Italy	2010	KP997816
A2	Italy	2010	KP997864
A2	Argentina	2011	KJ810923
A2	Italy	2011	KP997974
A2	Argentina	2012	KJ843188
A2	Italy	2012	KP997321
A2	Panama	2012	KP718102
A2	Panama	2012	KP718087
A2	Panama	2012	KP718070
A2	Argentina	2013	KJ843218
A2	Argentina	2013	KJ810926
A2	Cameroon	2013	KT749832
A2	Panama	2013	KP718080
A2	Uruguay	2013	KJ586809
A2	CapeVerde	2014	MF772344
A2	Bangladesh	2015	MF925386
A2	CapeVerde	2015	MF772350
A2	CapeVerde	2015	MF772349
A2	CapeVerde	2015	MF772348
A2	CapeVerde	2015	MF772347
A2	CapeVerde	2016	MF772346
A2	CapeVerde	2016	MF772345
A2	Kenya	2018	MK127847
A2	Germany	1992_1997	DQ788725
A2	Germany	1992_1997	DQ788726
A2	Germany	1995_1999	AF143303
A2	Martinique	2001_2004	HE974364
A2	Belgium	2003_2004	EU85990

D3	Sweden	1975	JX898688
D3	Sweden	1975	JX898686
D3	Canada	1984	GQ922001
D3	France	1990	AJ344117
D3	Canada	1995	GQ922002
D3	Sweden	1996	JX898691
D3	Sweden	1996	JX898690
D3	Sweden	1996	JX898693
D3	Russia	1997	JX090661
D3	Belgium	1998	FJ349209
D3	Belgium	1999	FJ349214
D3	Brazil	2003	KP090181
D3	Brazil	2003	KP090180
D3	Brazil	2004	MH724230
D3	Brazil	2004	MH724227
D3	Brazil	2004	MH724229
D3	Brazil	2004	MH724235
D3	Brazil	2004	MH724220
D3	Brazil	2004	MH724232
D3	USA	2004	KY810011
D3	Australia	2005	GQ486204
D3	Canada	2005	GQ486486
D3	Germany	2005	GQ486374
D3	Germany	2005	GQ486102
D3	SouthAfrica	2005	KM375775
D3	SouthAfrica	2005	KM375770
D3	France	2006	GQ486726
D3	France	2006	GQ486298
D3	Greece	2006	GQ486068
D3	Greece	2006	GQ486099
D3	Haiti	2006	FJ692506
D3	Poland	2006	GQ486073
D3	USA	2006	JN604313
D3	Brazil	2007	KY810013
D3	Brazil	2007	KY810040
D3	Brazil	2007	KY809962
D3	Brazil	2007	KY810023
D3	Brazil	2007	KY809990
D3	Brazil	2007	KY809995
D3	Brazil	2007	KY809958
D3	Brazil	2007	KY810044
D3	Brazil	2007	KY809969
D3	France	2007	GQ486759
D3	Germany	2007	GQ486698
D3	Serbia	2007	HQ236015
D3	SouthAfrica	2007	KT347090
D3	Spain	2007	GQ486678
D3	USA	2007	JN604301
D3	Argentina	2008	JN688679
D3	Brazil	2008	KY809965
D3	Brazil	2008	KY809964
D3	Brazil	2008	KY809976
D3	Brazil	2008	KY809948
D3	Brazil	2008	KY809981
D3	Brazil	2008	KY810006
D3	Brazil	2008	KY809982
D3	Brazil	2008	KY809966
D3	Sweden	2008	JX898699
D3	USA	2008	JN604170
D3	Botswana	2009	MH464854
D3	Botswana	2009	MH464843
D3	Botswana	2009	MH464838
D3	Botswana	2009	MH464848
D3	Brazil	2009	KY809950
D3	Brazil	2009	JN983941
D3	Brazil	2009	KY809963
D3	Italy	2009	KT288351
D3	Italy	2009	KP997348
D3	Italy	2009	KP997788
D3	Italy	2009	KP997759
D3	Argentina	2010	JN688708
D3	Italy	2010	KT275283
D3	Italy	2010	KP997882
D3	Italy	2010	KP997851
D3	Italy	2010	KP997891
D3	Italy	2010	KT275282
D3	Italy	2010	KP997847
D3	Italy	2010	KP997343
D3	Syria	2010	JN257181
D3	Uruguay	2010	KJ586811
D3	USA	2010	KX827292
D3	Argentina	2011	KJ843187
D3	Argentina	2011	JN688722
D3	Belarus	2011	KY816287
D3	Brazil	2011	KU847602
D3	Brazil	2011	KU847592
D3	Brazil	2011	KU847510
D3	Brazil	2011	KU847561
D3	France	2011	KP712811
D3	France	2011	KP712832
D3	Argentina	2012	KJ810915
D3	Brazil	2012	MH724219
D3	Brazil	2013	MH724222
D3	Cuba	2013	KM606703
D3	India	2013	KP322602
D3	Botswana	2015	MF979164
D3	Botswana	2015	MF979172
D3	Italy	2016	MG585269
D3	USA	2017	MF967563
D3	Brazil	2005_2007	JF815651
D3	Brazil	2005_2007	JF815633
D3	Brazil	2005_2007	JF815634
D3	Brazil	2005_2007	JF815627
D3	Brazil	2005_2007	JF815610
D3	Brazil	2005_2007	JF815632
D3	Brazil	2005_2007	JF815662
D3	Brazil	2005_2007	JF815637
D3	Brazil	2005_2007	JF815656
D3	Brazil	2005_2007	JF815622
D3	Brazil	2005_2007	JF815628
D3	Brazil	2005_2007	JF815621
D3	India	2009_2010	JN664909
D3	India	2009_2012	KC875319
D3	India	2009_2012	KC875332
D3	India	2009_2012	KC875321
D3	India	2009_2012	KC875318
D3	India	2009_2012	KC875315
D3	India	2009_2012	KC875324
D3	India	2009_2012	KC875328
D3	India	2009_2012	KC875325
D3	India	2009_2012	KC875316
D3	India	2009_2012	KC875314
D3	India	2009_2012	KC875331

D4	Canada	1984	GQ922003
D4	Canada	1984	GQ922005
D4	Brazil	2004	MH724251
D4	Kiribati	2004	HQ700537
D4	Canada	2005	GQ486255
D4	Canada	2005	GQ486265
D4	Germany	2005	GQ486604
D4	USA	2005	GQ486655
D4	Canada	2006	GQ486237
D4	Cuba	2006	KM606752
D4	Cuba	2006	KM606750
D4	Haiti	2006	FJ692536
D4	Haiti	2006	FJ692533
D4	USA	2006	JN604174
D4	Brazil	2007	KY810002
D4	France	2007	GQ486699
D4	Brazil	2008	KJ470891
D4	Brazil	2008	KJ470897
D4	Brazil	2009	JN983920
D4	Brazil	2009	JN983921
D4	Brazil	2009	JF298901
D4	Brazil	2009	JN983918
D4	Brazil	2009	JN983924
D4	Brazil	2010	KJ470884
D4	Brazil	2010	JN983933
D4	Brazil	2010	JN983932
D4	Brazil	2010	KJ470898
D4	Brazil	2010	KJ470885
D4	Brazil	2010	JN983931
D4	Brazil	2011	KU847535
D4	Brazil	2011	KU847624
D4	Brazil	2011	KU847488
D4	Brazil	2011	KU847552
D4	France	2011	KP712825
D4	France	2011	KP712859
D4	Belarus	2012	KY816288
D4	Brazil	2012	KU847741
D4	Brazil	2012	KX302100
D4	Brazil	2012	KX302092
D4	Brazil	2012	KX302111
D4	Brazil	2012	KX302088
D4	Brazil	2012	KX302089
D4	Kenya	2012	KP168416
D4	SouthAfrica	2012	MH481861
D4	SouthAfrica	2012	MH481862
D4	Brazil	2013	KX302126
D4	Brazil	2013	KX302124
D4	Brazil	2013	KX302128
D4	Brazil	2013	KX302114
D4	Brazil	2013	KX302127
D4	USA	1990_1995	A08967
D4	Martinique	2001_2004	HE974372
D4	Martinique	2001_2004	HE974382
D4	Brazil	2005_2007	JF815648
D4	India	2010_2012	KF192839
D4	India	2010_2012	KF192840
D4	India	2010_2012	KF192841
E	Angola	2007	KF849713
E	Angola	2007	KF849726
E	Belgium	1995	FJ349226
E	Botswana	2007	KR139749
E	Botswana	2009	MH464856
E	Cameroon	1994	AB194947
E	Cameroon	2005	MK174151
E	Cameroon	2005	FN545827
E	Cameroon	2010	MK174160
E	Cameroon	2011	MH580647
E	Cameroon	2012	KY494109
E	Cameroon	2013	KY493952
E	Cameroon	2014	KY494008
E	Cameroon	2015	KY494167
E	Cameroon	2017	MH165306
E	Cameroon	2006_2007	FN545842
E	CapeVerde	2015	MF772360
E	CapeVerde	2016	MF772362
E	Colombia	2007	JQ023665
E	Cuba	2006	KM606743
E	Cuba	2013	KM606738
E	Egipt	2010	KU736892
E	Ethiopia	2010	KU736893
E	France	2006	GQ486093
E	France	2006	GQ486656
E	France	2011	KP712861
E	Germany	2005	GQ486548
E	Ghana	2000	AB205189
E	Ghana	2006	MK174182
E	Ghana	2007	MK174156
E	Ghana	2009	MK174163
E	Ghana	2013	KU711647
E	Ghana	2014	KU711641
E	Guinea	2005	MK174185
E	Guinea	2006	GQ161829
E	Guinea	2013	KX186584
E	Japan	1990_2002	AP007262
E	Liberia	2005	MK174186
E	Liberia	2009	KU736895
E	Liberia	2010	KU736894
E	Martinique	2001_2004	HE974380
E	Mexico	2011	KT192626
E	Netherlands	2005	GQ486411
E	Netherlands	2015	MG383615
E	Nigeria	2005	MH918655
E	Nigeria	2007	HM363573
E	Norway	2007	MK174157
E	RepCongo	2013_2014	MH253773
E	SaudiArabia	2008	AB915175
E	Senegal	2008	MK174158
E	Somalia	2007	KU736900
E	SouthAfrica	2006	KM374874
E	SouthAfrica	2012	KF922439
E	Sudan	2008	KF170742
E	Sudan	2010	KU736909
E	Sudan	2011	KM108622
E	Thailand	2008	GQ161792
E	UnitedKingdom	2018	MK321264
E	USA	2008	JN604166

ANNEX 3

Part A

HCV RNA extraction, amplification, and sequencing of partial NS5B genomic region (Okamoto region).

HCV RNA extraction (QIAamp Viral RNA Mini kit Kit, QIAGen, Werfen)

c) Reagent's preparation when opening the kit for the first time

- Add 310 µl Buffer AVE to the tube containing 310 µg lyophilized carrier RNA (contained in the kit) to obtain a solution of 1 µg/µl. Dissolve the carrier RNA thoroughly, divide it into conveniently sized aliquots, and store it at –20°C.

- Buffer AW1: add 130 ml of 96-100% ethanol to 98 ml of concentrated Buffer AW1 and mix by inversion 5-10 times.

- Buffer AW2: pipette 160 ml of 96-100% ethanol into 66 ml of buffer AW2 and mix by inversion 5-10 times.

- Note: Unused volume of reconstituted Buffer AW1 and Buffer AW2 can be stored at room temperature for 1 year.

d) HCV RNA extraction

- Shake AVL buffer before using it and if there is any precipitate, dissolve it in a bath at 80ºC. Do not heat for more than 5 minutes

- Prepare 560 µl of AVL plus 5,6 µl of carrier RNA for each sample immediately before extraction.

- Prepare 140 µl of negative extraction control from a pool of negative plasmas for this virus.

- Pipette 140 µl of plasma sample and positive and negative controls into 1.5ml microtubes.

- Add 560 µl of prepared AVL Buffer containing carrier RNA into the previous microcentrifuge tube.

- Vortex intermittently for 15 seconds; to promote an efficient lysis it is essential that the sample and the AVL buffer are mixed until a homogeneous solution is obtained.

- Incubate at room temperature (15-25°C) for 10 min.

- Briefly centrifuge the tube to remove drops from the inside of the lid.

- Add 560 µl of ethanol (96–100%) to the sample and Vortex intermittently for 15 seconds.

- Briefly centrifuge the tube to remove drops from inside the lid.

- Apply 630 µl of the mixture above into a QIAamp Mini column (in a 2 ml collection tube) without wetting the rim.

- Centrifuge at 6,000g for 1 min. Place the column into a clean 2 ml collection tube and discard the tube containing the filtrate.

- Open the column lid and repeat the two previous steps.

- Open the column lid, add 500 µl of Buffer AW1 and centrifuge at 6,000g for 1 min. Place the column in a clean 2 ml collection tube and discard the tube containing the filtrate.

- Open the column lid, add 500 µl of Buffer AW2 and centrifuge at full speed (20,000g) for 3 min.

- Place the column in a new 2 ml collection tube, discard the old collection tube with the filtrate and centrifuge at full speed for 1 min.

- Place the column in a clean 1.5 ml microcentrifuge tube, discard the old collection tube containing the filtrate.

- Put the column into a 1.5ml microtube and add 60 µl of Buffer AVE equilibrated to room temperature.

- Close the cap and incubate at room temperature for 3 min. Centrifuge at 6,000g for 1 min and store at -80ºC until use or proceed to amplification step immediately.

<u>Partial amplification of HCV NS5b gene_Okamoto region</u> (Titan One Tube RT-PCR System; Roche Diagnostics, Switzerland)

Note: The PCR reaction is performed with RNase out (Themofisher Scientific, USA) which is not included in this kit. Titan enzyme and RNase Out should only be removed from the freezer at the time of use.

- Put the reagents of amplification kit at room temperature, shake them slightly, spin them and place them on ice.

  Note: the enzyme should only be removed from the freezer on the moment it is going to be used.

- Mark 0.2 ml microtubes for each of the samples, for the positive control, for the negative control of extraction and for the negative control of PCR.

- Prepare the 2 Titan mixtures for the PCR reaction in 2 different 1.5mL microtubes.

PCR mixture 1 (for 1 sample)

dNTPs (100mM) - 0,2 µl

Primer OKA1 (10µM) - 1 µl

Primer OKA2 (10µM) - 1 µl

DTT (100mM) - 1,25 µl

Nuclease free water - 3,55 µl

RNase Out (40U/µl) - 0,5 µl

Total Volume - 7,5 µl

PCR mixture 2 (for 1 sample)

Titan RT-PCR Reaction Buffer (5X) - 5 µl

Nuclease free water - 7 µl

Titan Enzyme - 0,5 µl

Total Volume - 12,5 µl

- Homogenize and distribute 7.5 µl of mix 1 + 12.5 µl of mix 2 in each 0.2ml tube.
- Add 5 µl of extracted RNA and nuclease free water (PCR negative control) to the respective tubes.
- Place the tubes in a thermal cycler under the following cycles:

55ºC	30 min	
94ºC	2 min	
94ºC	1 min	
59ºC	30 secs	10 cycles
68ºC	45 secs	
94ºC	10 secs	
59ºC	30 secs	30 cycles
68ºC	45 secs (+5 secs/cycle)	
68ºC	7 min	
4ºC	∞	

- After the end of the PCR reaction, store the amplified products at -20ºC or proceed with the next step.

Agarose gel electrophoresis

- Prepare a 2.0% agarose gel.
- Fill an electrophoresis cuvette with 1x TAE buffer.
- Prepare the samples to be applied to the gel, adding 10µl of each to 2µl of Orange G dye.
- To 0.5µl of the molecular weight marker GeneRuler 100 bp DNA Ladder, 0,5 µg/µl (ThermoFisher Scientific), 2µl of the dye and 7.5µl of water are added to make a volume of 10µl.

- Apply the stained samples into the wells of the agarose gel and apply electric current (70 Volts) for approximately 30 minutes.
- Observe the gel in a transilluminator where a 372 bp band should be visualized for samples and positive control.

PCR products purification using a spin column system (JetQuick PCR Product Purification Spin Kit; Genomed, Germany)

- Place a 1.5 ml microtube with nuclease free water to heat at 65ºC.
- Add 60µl of H1 kit solution to remaining 15µl of 2nd round PCR product.
- Mix and transfer the mixture to a purification column previously inserted in a supernatant collection tube.
- Centrifuge at 15000g for 1 minute.
- Discard the contents of the collection tube.
- Add 500µl of H2 solution to the column.
- Repeat the above centrifugation and discard steps.
- Centrifuge at 16000g for 1 minute to completely dry the column membrane.
- Insert the column into a 1.5ml microtube.
- Add 20-35µl of nuclease free water at 65ºC to each column.
  Note: Add a higher volume, up to 50µl, if the band obtained after electrophoresis is too strong.
- Wait 1 minute.
- Centrifuge at 16000g for 2 minutes.
- Store purified products at -20°C or proceed with the sequencing reaction.

Sequencing Reaction (BigDye terminator v.3.1 Cycle Sequencing Kit; Applied Biosystems)

- Thaw the reagents needed for the sequencing reactions and the purified PCR products:
    a. BigDye terminator v.3.1 (Applied Biosystems)
    b. BigDye terminator 5X sequencing buffer (Applied Biosystems)

     c. Primers Forward and Reverse from the nested PCR (5µM)

     d. Nuclease free water

- Place the thawed reagents on ice and prepare a mixture to each primers as described below:

Sequencing mixture (for 1 sample)

BigDye terminator v.3.1 - 0.5 µl

Primer (5µM) - 1 µl

BigDye Buffer (5X) - 1.75 µl

- Mix the reagents by pipetting up and down several times and make a spin to the microtube.
- Distribute 3.25 µl of each mix into the two different tubes.
- Add between 0.3 - 1 µl of the previously purified PCR product to each tube, according to the intensity of the band observed on the agarose gel.
- Complete the reaction volume to 10 µl with nuclease free water and place the samples in the thermocycler under the following conditions:

Sequencing cycle

96ºC	1 min	
96ºC	10 secs	
56ºC	5 secs	25 cycles
60ºC	4 min	
4ºC	7 min	
4ºC	∞	

Sequencing products purification through Ethanol/Sodium Acetate

- Prepare a sodium acetate/ethanol solution by mixing 2 µl of 3M sodium acetate (pH 4.6) with 50 µl of 100% ethanol, per reaction.

- Add 52 µl of the previous solution to each sequencing reaction, transfer to a 96 well plate and cover the plate with aluminium foil.

- Vortex the plate to mix.

- Centrifuge the plate at 2000g for 20 minutes.

- Remove the aluminium foil carefully trying to not disturb the sediment.

- Immediately place several absorbent paper towels without residues on top of the plate and invert it.

- Centrifuge the inverted plate at 150g for 1 minute.

- Add 150 µl of 70% ethanol, prepared previously, and cover the plate with a new aluminium foil.

- Centrifuge at 2000g for 5 minutes.

- Remove again the aluminium foil carefully trying to not disturb the sediment.

- Invert the plate over clean towels and centrifuge the inverted plate at 150g for 1 minute.

- After centrifuging, allow the plate to air dry for approximately 1 minute to eliminate any remaining ethanol.

- Cover the plate with aluminium foil to store between –15ºC and –25ºC in the dark, but no longer than 2 weeks or add 20 µl of Hi-Di^{TM} Formamide (Applied Biosystems) and move to the processing phase of the sequences in the ABI Prism 3100-Avant Genetic Analyzer according with the equipment manual.

Part B

HCV RNA extraction, amplification, and sequencing of complete NS5 genomic region.

<u>HCV RNA extraction</u> (NucliSENS® EasyMAG® system v2.0 - with silica; Boseind, Boxtel, The Netherlands)

a) Workflow in the automatic system.

- Enter sample set data in the computer associated to the equipment.

- Set samples matrix, sample size, and choose 45 µl for elution volume.

- Load 500 µl of lysed plasma samples into 8-well sample vessels, previously positioned in the vessel's carrier.

- Insert aspiration tips sets, sample vessel strips and scan bar code.

- EasyMag automatically will verify samples, tip insertion and on-board reagents.

- Add silica.

- Start run that will end about 1 hour later.

Note: Always introduce a positive and a negative known samples as positive and negative control of extraction, respectively.

b) Principle of the procedure inside the equipment

- During the run all the target nucleic acid is captured by magnetic silica particles which will be attracted by the EasyMag magnetic device

- The system will purify the nucleic acids through several washing steps.

- Heating steps releases nucleic acids from silica.

- During the elution (and last) step silica will be separated from the eluted by the magnetic device.

Amplification of HCV NS5A and NS5B genomic region

a) First round PCR (QIAGEN OneStep RT-PCR Kit; Hilden, Germany)

- Put the reagents of amplification kit at room temperature, shake them slightly, spin them and place them on ice.

    Note: the enzyme should only be removed from the freezer on the moment it is going to be used.

- Mark 0.2 ml microtubes for each of the samples, for the positive control, for the negative control of extraction and for the negative control of PCR.

- Prepare the mixture for the PCR reaction in a 1.5mL microtube.

PCR mixture (for 1 sample)

OneStep RT-PCR Buffer (5x) - 10µL

dNTPs Mix (10mM each) – 2µL

Primer FW1_PCR (10µM) – 2.5µL

Primer RV1_PCR (10µM) – 2.5µL

OneStep RT-PCR Enzyme Mix – 2.5µL

Protector RNase inhibitor (40U/µL) – 0.25µL

Nuclease free water – 20.25µL

- Homogenize and distribute 40µl of mix in each 0.2ml tube.

- Add 10µl of extracted RNA and nuclease free water (PCR negative control) to the respective tubes.

- Place the tubes in a thermal cycler under the following cycles:

45ºC	30 min	
95ºC	15 min	
94ºC	10 secs	
63ºC	45 secs (-1 ˚C/cycle)	16 cycles
68ºC	5 min	
94ºC	10 secs	
48ºC	45 secs	30 cycles
68ºC	5 min (+3 secs/cycle)	
68ºC	10 min	
4ºC	∞	

b) After the end of the PCR reaction, store the amplified products at -20ºC or proceed with the second round PCR (Expand High Fidelity PCR System; Roche, Switzerland)

PCR mixture 1 (for 1 sample)

dNTPs (100mM) – 0.4µl

Primer FW2_PCR (10µM) – 1.5µl

Primer RV2_PCR (10µM) – 1.5µl

Nuclease free water – 19.6µl

Total Volume - 23µl

PCR mixture 2 (for 1 sample)

Expand High Fidelity Buffer without $MgCl_2$ (10X) - 5µl

$MgCl_2$ (25mM) - 6µl

DMSO (100%) – 0.2µl

Nuclease free water – 12.93µl

Expand High Fidelity enzyme mix (3.5U/µl) – 0.87µl

Total Volume - 25µl

- Distribute 23 µL of mix 1 into each 0.2mL microtubes.
- Add 2 µL of 1st round PCR DNA products to the 23µL of mix 1.
- Add 25 µL of Mix 2 to make a final reaction volume of 50 µL.
- Place the tubes in a thermal cycler under the following cycles:

95ºC	3 min	
94ºC	15 secs	
60ºC	30 secs (−1 ˚C/cycle)	16 cycles
68ºC	3 min	
94ºC	10 secs	
45ºC	30 secs	30 cycles
68ºC	3 min (+5 secs/cycle)	
68ºC	7 min	
4ºC	∞	

Details about primers used in amplification reactions are available in table A3 at the end of this annex.

- After the end of the PCR reaction, store the amplified products at -20ºC or proceed with the next step.

Agarose gel electrophoresis

- Prepare a 1.0% agarose gel.
- Fill an electrophoresis cuvette with 1x TAE buffer.
- Prepare the samples to be applied to the gel, adding 8µl of each to 2µl of blue juice loading buffer (Thermofisher Scientific, USA)
- To 1µl of the molecular weight marker GeneRuler 1Kb DNA Ladder, 0,5 µg/µl; (ThermoFisher Scientific, USA), 1µl of the loading dye provided with the ladder and 4µl of water are added to make a volume of 6µl.

- Apply the stained samples into the wells of the agarose gel and apply electric current (70 Volts) for approximately 60 minutes.
- Observe the gel in a transilluminator where a 3700bp band should be visualized for samples and positive control.

PCR products purification using (ExoSAP-IT™ PCR Product Cleanup; Applied Biosystems, USA)

- Remove ExoSAP-IT™ reagent from –20°C freezer and keep on ice throughout this procedure.
- Mix 5 µL of PCR reaction product with 2 µL of ExoSAPIT™ reagent for a combined 7 µL reaction volume.
- Note: When treating PCR product volumes greater than 5 µL, simply increase the amount of ExoSAP-IT™ reagent proportionally.
- Incubate at 37°C for 15 minutes to degrade remaining primers and nucleotides.
- Incubate at 80°C for 15 minutes to inactivate ExoSAP-IT™ reagent.
- Store purified PCR products at –20°C until required or proceed with the sequencing reaction

Sequencing of the complete NS5 genomic region

The sequencing process was separated in two sets of mix reactions, one for genomic regions of difficult hybridization by the designed sequencing primers (Sequencing Mix A) and other for regions of easy hybridization (Sequencing Mix B).

- Thaw the reagents needed for the sequencing reactions and the purified PCR products:
    a) BigDye terminator v.3.1 (Applied Biosystems)
    b) BigDye terminator 5X sequencing buffer (Applied Biosystems)
    c) Primers Forward and Reverse from the nested PCR (5µM)
    d) Nuclease free water

- Place the thawed reagents on ice and prepare a mixture to each primer as described below:

Sequencing mixture A (for 1 sample)

BigDye terminator v.3.1 - 2µl

Primer (5µM) - 1µl

Buffer (5X) - 3µl

Nuclease free water - 13 µl

Note: Mix A is for primers RV4, FW6, and RV9 (Table below). If needed, to complete the NS5 sequencing product apply same reaction conditions to sequencing back-up primers: FW2PCR, RV1, FW2, RV2, FW3, RV3, RV5, FW7, 1bRV8, 1aFW9, 1aFW10 and 1bFW11.

- Mix the reagents by pipetting up and down several times and make a spin to the microtube.
- Distribute 19µl of each primer mix into the different tubes.
- Add 1 µl of PCR product purified.

Sequencing mixture B (for 1 sample)

BigDye terminator v.3.1 - 1µl

Primer (5µM) - 1µl

Buffer (5X) – 1.5µl

Nuclease free water – 5.5 µl

Note: Mix A is for primers RV2PCR, FW1, FW4, FW5, RV6, 1bRV7 and 1aFW8.

- Mix the reagents by pipetting up and down several times and make a spin to the microtube.

- Distribute 9µl of each primer mix into the different tubes.

- Add 1 µl of PCR product purified.

- Place the sequencing reactions in the thermocycler under the following conditions (both for mix A and B):

<u>Sequencing cycle</u>

96ºC	5 min	
94ºC	5 secs	
50ºC	10 secs	35 cycles
60ºC	4 min	
4ºC	∞	

Sequencing products purification through Ethanol/Sodium Acetate

- Prepare a sodium acetate/ethanol solution by mixing 2 µl of 3M sodium acetate (pH 4.6) with 50 µl of 100% ethanol, per reaction.

- Add 52 µl of the previous solution to each sequencing reaction, transfer to a 96 well plate and cover the plate with aluminium foil.

- Vortex the plate to mix.

- Centrifuge the plate at 2000g for 20 minutes.

- Remove the aluminium foil carefully trying to not disturb the sediment.

- Immediately place several absorbent paper towels without residues on top of the plate and invert it.

- Centrifuge the inverted plate at 150g for 1 minute.

- Add 150 µl of 70% ethanol, prepared previously, and cover the plate with a new aluminum foil.

- Centrifuge at 2000g for 5 minutes.

- Remove again the aluminium foil carefully trying to not disturb the sediment.

- Invert the plate over clean towels and centrifuge the inverted plate at 150g for 1 minute.

- After centrifuging, allow the plate to air dry for approximately 1 minute to eliminate any remaining ethanol.

- Cover the plate with aluminium foil to store between –15ºC and –25ºC in the dark, but no longer than 2 weeks or add 20 µl of Hi-Di^{TM} Formamide (Applied Biosystems) and move to the processing phase of the sequences in the ABI PRISM 3130xl Genetic Analyzer (ThermoFisher Scientific, USA) according with the equipment manual.

Table A3 Primers details used in PCR and sequencing reactions.

Primer name	Purpose	Primer sequence (5'- 3')	Location (in H77) (bp)	Average Tm
FW1PCR	RT-PCR NS5A-NS5B region	GTCGTCACBAGYACCTGG	5311 - 5328	56ºC
RV1PCR	RT-PCR NS5A-NS5B region	GAGACASGCTGTGATAWATG	9297 - 9316	51ºC
FW2PCR	Nested PCR NS5A-NS5B region/SequencingNS5A-NS5B region (Back-up option)	GTGGTCATHGT GGYAGG	5386 - 5403	53ºC
RV2PCR	Nested PCR NS5A-NS5B region/SequencingNS5A-NS5B region	CCCTATTGATYTCACCTGG	9055 - 9073	51ºC
FW1	Sequencing NS5A-NS5B region	GCHGTGCARTGGATGAA	6118 - 6134	53ºC
RV1	Sequencing NS5A-NS5B Region (Back-up option)	CCRTTYYTGACATGTCC	6493 - 6509	49ºC
FW2	Sequencing NS5A-NS5B Region (Back-up option)	TGGMGRCARG AGATG	7093 - 7107	49ºC
RV2	Sequencing NS5A-NS5B Region (Back-up option)	GGRTCRGTRAGCATGGA	6901 - 6917	53ºC
FW3	Sequencing NS5A-NS5B Region (Back-up option)	TCYTCYATG CCCCCYCT	7559 - 7574	57ºC
RV3	Sequencing NS5A-NS5B Region (Back-up option)	TTRTTYTCYG ACTCMAC	7126 - 7142	46ºC
FW4	Sequencing NS5A-NS5B region	AARGTCACHTTTGACAG	7816 - 7832	46ºC
RV4	Sequencing NS5A-NS5B region	GCBGARTGYGGGGGCGTCAG	7936 - 7955	65ºC
FW5	Sequencing NS5A-NS5B region	AAGCCAGCTCGYCTYATCGT	8128 - 8147	59ºC
RV5	Sequencing NS5A-NS5B Region (Back-up option)	ACGAGCATBGTGCAGTC	8593 - 8609	55ºC
FW6	Sequencing NS5A-NS5B region	CTTCACGGAGGCTATGAC	8679 - 8696	52ºC
RV6	Sequencing NS5A-NS5B region	GACCADGAYCCGTCRCT	7606 - 7619	55ºC
FW7	Sequencing NS5A-NS5B Region (Back-up option)	TYTACCARTGYTGTGAC	8381 - 8397	48ºC
1bRV7	Sequencing GT1b NS5A-NS5B region	CATCTCCTG CCG CCACA	7067 - 7083	58ºC
1aFW8	Sequencing GT1a NS5A-NS5B region	ACACTCG CTG CCVCTGTG	6433 - 6450	60ºC
1bRV8	Sequencing GT1b NS5A-NS5B Region (Back-up option)	AGGTCAAGTGG CTCAATGGA	8987 - 9004	56ºC
1aFW9	Sequencing GT1a NS5A-NS5B Region (Back-up option)	CTTCACGGAGGCTATGACC	8636 - 8654	55ºC
RV9	Sequencing NS5A-NS5B region	ACGATRAGRCGAGCTGGCTT	8083 - 8102	59ºC
1aFW10	Sequencing GT1a NS5A-NS5B Region (Back-up option)	GACAGCAAGACACACTCC	8816 - 8833	53ºC
1bFW11	Sequencing GT1b NS5A-NS5B Region (Back-up option)	AACTCCTGG CTAGGCAACAT	8848 - 8867	56ºC

192

ANNEX 4

Figure A4. Complete HBV MCC tree without the collapsed branches

(see next page)

Branches Color Code

Portugal	Brazil	Cape Verde
Botswana	Italy	Spain
USA	France	India
Vietnam	Nigeria	Lybia
Cameroon	Kenya	Mali
Haiti	Rwanda	South Africa
Zimbabwe	Algeria	Gaboon
Japan	Belgium	Congo Democ Rep
Chad	China	Caribean Sea
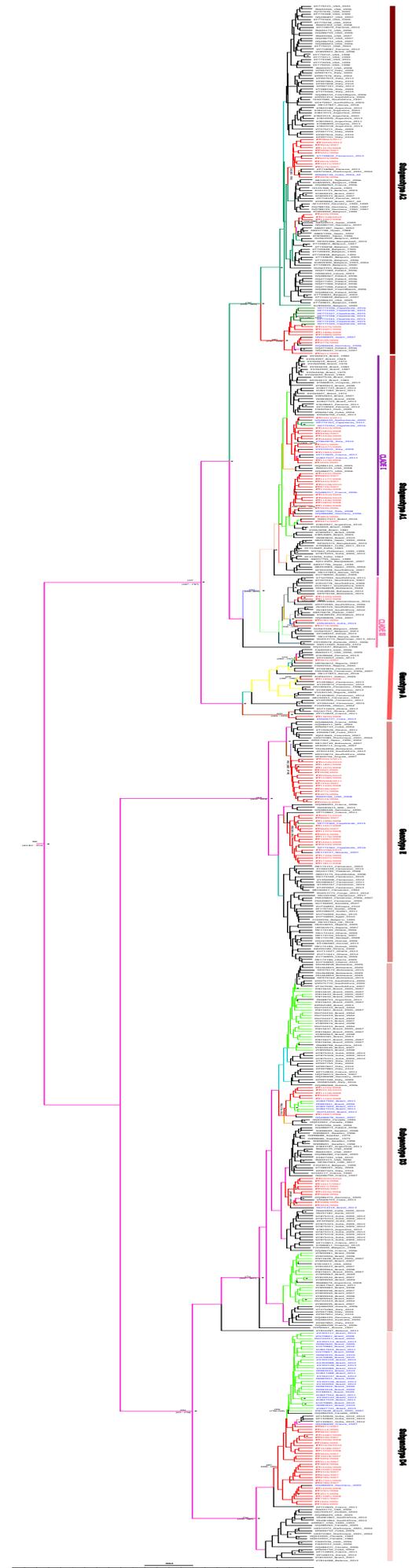
ANNEX 5

Table A5. HCV sequences with inconsistent clade classification after comparison of the phylogenetic analysis and the geno2pheno[HCV]

Sequences ID	Clade by phylogenetic analysis	Clade by geno2pheno[HCV]
10154.1a_clade2	2	1
10509.1a_clade2	2	1
10953.1a_clade2	2	1
11296.1a_clade2	2	1
11519.1a_clade2	2	1
11528.1a_clade2	2	1
11812.1a_clade2	2	1
11869.1a_clade2	2	1
11929.1a_clade2	2	1
12710.1a_clade2	2	1
13096.1a_clade2	2	1
13211.1a_clade2	2	1
13515.1a_clade2	2	1
13602.1a_clade2	2	1
13701.1a_clade2	2	1
14971.1a_clade2	2	1
14973.1a_clade2	2	1
15038.1a_clade2	2	1
15249.1a_clade2	2	1
15637.1a_clade2	2	1
15644.1a_clade2	2	1
15682.1a_clade2	2	1
15710.1a_clade2	2	1
15941.1a_clade2	2	1
16342.1a_clade2	2	1

ANNEX 6

Table A6. Comparison of genotypes/subtypes/clades obtained through different methods for the 58 HCV/1a sequences; Sequences classified as clade 1 are colored blue to better distinguish themselves from clade 2 in the table.

	HCV Genotyping/Subtyping Methods		
Sample Code	**InnoLiPA**	**Geno2Pheno**	**ML Phylogenetic Tree**
PT 102985	1a	1a/clade II	1a/clade II
PT 107537	1a	1a/clade I	1a/clade I
PT 108200	1a	1a/clade II	1a/clade II
PT 108338	1a	1a/clade I	1a/clade I
PT 109872	1a	1a/clade I	1a/clade I
PT 110257	1a	1a/clade I	1a/clade I
PT 110270	1a	1a/clade II	1a/clade II
PT 110368	1a	1a/clade II	1a/clade II
PT 110779	1a	1a/clade II	1a/clade II
PT 111766	1a	1a/clade II	1a/clade II
PT 113509	1a	1a/clade II	1a/clade II
PT 113563	1a	1a/clade II	1a/clade II
PT 113894	1a	1a/clade II	1a/clade II
PT 113968	1a	1a/clade I	1a/clade I
PT 114606	1a	1a/clade II	1a/clade II
PT 114771	1a	1a/clade II	1a/clade II
PT 114784	Inconclusive	1a/clade I	1a/clade I
PT 114989	1a	1a/clade II	1a/clade II
PT 114999	1a	1a/clade I	1a/clade I
PT 115059	1a	1a/clade II	1a/clade II
PT 115669	1a	1a/clade II	1a/clade II
PT 116188	1a	1a/clade II	1a/clade II
PT 116276	1a	1a/clade II	1a/clade II
PT 116604	1a	1a/clade I	1a/clade I
PT 116828	1a	1a/clade II	1a/clade II
PT 117188	1a	1a/clade II	1a/clade II
PT 117203	1a	1a/clade I	1a/clade I
PT 117272	1a	1a/clade II	1a/clade II
PT 118129	1b	1a/clade II	1a/clade II
PT 118658	1a	1a/clade II	1a/clade II
PT 120956	1a	1a/clade II	1a/clade II
PT 122802	1a	1a/clade II	1a/clade II
PT 123017	1a	1a/clade II	1a/clade II
PT 123018	1a	1a/clade I	1a/clade I
PT 123019	1a	1a/clade II	1a/clade II
PT 123020	1a	1a/clade II	1a/clade II
PT 123445	1a	1a/clade I	1a/clade I
PT 123631	1a	1a/clade II	1a/clade II
PT 123728	1a	1a/clade II	1a/clade II
PT 124028	1a	1a/clade II	1a/clade II
PT 124060	1a	1a/clade II	1a/clade II
PT 124736	1a	1a/clade II	1a/clade II
PT 125544	1a	1a/clade II	1a/clade II
PT 125662	1a	1a/clade II	1a/clade II
PT 125778	1a	1a/clade I	1a/clade I
PT 125931	1a	1a/clade II	1a/clade II
PT 126219	1a	1a/clade II	1a/clade II
PT 126401	1a	1a/clade II	1a/clade II
PT 127160	1a	1a/clade II	1a/clade II
PT 127831	1a	1a/clade I	1a/clade I
PT 130114	1a	1a/clade II	1a/clade II
PT 131110	1a	1a/clade II	1a/clade II
PT 131663	1a	1a/clade II	1a/clade II
PT 131865	1a	1a/clade I	1a/clade I
PT 132010	1a	1a/clade II	1a/clade II
PT 139290	1a	1a/clade II	1a/clade II
PT 144030	1a	1a/clade II	1a/clade II
PT 149353	1a	1a/clade I	1a/clade I

ANNEX 7

ISO Standard Country Codes

Country	Code
Afghanistan	AF
Aland Islands	AX
Albania	AL
Algeria	DZ
American Samoa (US)	AS
Andorra	AD
Angola	AO
Anguilla (UK)	AI
Antarctica	AQ
Antigua and Barbuda	AG
Argentina	AR
Armenia	AM
Aruba	AW
Australia	AU
Austria	AT
Azerbaijan	AZ
Bahamas	BS
Bahrain	BH
Bangladesh	BD
Barbados	BB

Country	Code
Belarus	BY
Belgium	BE
Belize	BZ
Benin	BJ
Bermuda (UK)	BM
Bhutan	BT
Bolivia	BO
Bonaire, Sint Eustatius, and Saba	BQ
Bosnia and Herzegovina	BA
Botswana	BW
Bouvet Island	BV
Brazil	BR
British Indian Ocean Territory	IO
British Virgin Islands (UK)	VG
Brunei Darussalam	BN
Bulgaria	BG
Burkina Faso	BF
Burundi	BI
Cambodia	KH
Cameroon	CM
Canada	CA
Cape Verde	CV

Country	Code
Cayman Islands (UK)	KY
Central African Republic	CF
Chad	TD
Chile	CL
China	CN
Christmas Island (AU)	CX
Cocos (Keeling) Islands (AU)	CC
Colombia	CO
Comoros	KM
Congo, Democratic Republic of the	CD
Congo, Republic of the	CG
Cook Islands (NZ)	CK
Costa Rica	CR
Côte D'Ivoire	CI
Croatia	HR
Cuba	CU
Curaçao	CW
Cyprus	CY
Czech Republic	CZ
Denmark	DK
Djibouti	DJ
Dominica	DM

Country	Code
Dominican Republic	DO
Ecuador	EC
Egypt	EG
El Salvador	SV
Equatorial Guinea	GQ
Eritrea	ER
Estonia	EE
Ethiopia	ET
Falkland Islands (UK)	FK
Faroe Islands (DK)	FO
Fiji	FJ
Finland	FI
France	FR
French Guiana (FR)	GF
French Polynesia (FR)	PF
French Southern Territories	TF
Gabon	GA
Gambia	GM
Georgia	GE
Germany	DE
Ghana	GH
Gibraltar (UK)	GI

Country	Code
Greece	GR
Greenland (DK)	GL
Grenada	GD
Guadeloupe (FR)	GP
Guam (US)	GU
Guatemala	GT
Guernsey	GG
Guinea	GN
Guinea-Bissau	GW
Guyana	GY
Haiti	HT
Heard Island and McDonald Islands	HM
Holy See (Vatican City)	VA
Honduras	HN
Hong Kong (CN)	HK
Hungary	HU
Iceland	IS
India	IN
Indonesia	ID
Iran	IR
Iraq	IQ
Ireland	IE

Country	Code
Isle of Man	IM
Israel	IL
Italy	IT
Jamaica	JM
Japan	JP
Jersey	JE
Jordan	JO
Kazakhstan	KZ
Kenya	KE
Kiribati	KI
Korea, Democratic People's Republic (North)	KP
Korea, Republic of (South)	KR
Kuwait	KW
Kyrgyzstan	KG
Laos	LA
Latvia	LV
Lebanon	LB
Lesotho	LS
Liberia	LR
Libya	LY
Liechtenstein	LI
Lithuania	LT

Country	Code
Luxembourg	LU
Macau (CN)	MO
Macedonia	MK
Madagascar	MG
Malawi	MW
Malaysia	MY
Maldives	MV
Mali	ML
Malta	MT
Marshall Islands	MH
Martinique (FR)	MQ
Mauritania	MR
Mauritius	MU
Mayotte (FR)	YT
Mexico	MX
Micronesia, Federated States of	FM
Moldova, Republic of	MD
Monaco	MC
Mongolia	MN
Montenegro	ME
Montserrat (UK)	MS
Morocco	MA

Country	Code
Mozambique	MZ
Myanmar	MM
Namibia	NA
Nauru	NR
Nepal	NP
Netherlands	NL
New Caledonia (FR)	NC
New Zealand	NZ
Nicaragua	NI
Niger	NE
Nigeria	NG
Niue	NU
Norfolk Island (AU)	NF
Northern Mariana Islands (US)	MP
Norway	NO
Oman	OM
Pakistan	PK
Palau	PW
Palestinian Territories	PS
Panama	PA
Papua New Guinea	PG
Paraguay	PY

Country	Code
Peru	PE
Philippines	PH
Pitcairn Islands (UK)	PN
Poland	PL
Portugal	PT
Puerto Rico (US)	PR
Qatar	QA
Reunion (FR)	RE
Romania	RO
Russia	RU
Rwanda	RW
Saint Barthelemy	BL
Saint Helena (UK)	SH
Saint Kitts and Nevis	KN
Saint Lucia	LC
Saint Martin (French Part)	MF
Saint Pierre & Miquelon (FR)	PM
Saint Vincent and the Grenadines	VC
Samoa	WS
San Marino	SM
Sao Tome and Principe	ST
Saudi Arabia	SA

Country	Code
Senegal	SN
Serbia	RS
Seychelles	SC
Sierra Leone	SL
Singapore	SG
Sint Maarten (Dutch Part)	SX
Slovakia	SK
Slovenia	SI
Solomon Islands	SB
Somalia	SO
South Africa	ZA
South Georgia & South Sandwich Islands (UK)	GS
South Sudan	SS
Spain	ES
Sri Lanka	LK
Sudan	SD
Suriname	SR
Svalbard and Jan Mayen	SJ
Swaziland	SZ
Sweden	SE
Switzerland	CH
Syria	SY

Country	Code
Taiwan	TW
Tajikistan	TJ
Tanzania	TZ
Thailand	TH
Timor-Leste	TL
Togo	TG
Tokelau	TK
Tonga	TO
Trinidad and Tobago	TT
Tunisia	TN
Turkey	TR
Turkmenistan	TM
Turks and Caicos Islands (UK)	TC
Tuvalu	TV
Uganda	UG
Ukraine	UA
United Arab Emirates	AE
United Kingdom	UK
United States	US
United States Minor Outlying Islands	UM
Uruguay	UY
Uzbekistan	UZ

Country	Code
Vanuatu	VU
Venezuela	VE
Vietnam	VN
Virgin Islands (US)	VI
Wallis and Futuna (FR)	WF
Western Sahara	EH
Yemen	YE
Zambia	ZM
Zimbabwe	ZW