This thesis has been submitted in fulfilment of the requirements for a postgraduate degree (e.g. PhD, MPhil, DClinPsychol) at the University of Edinburgh. Please note the following terms and conditions of use:

- This work is protected by copyright and other intellectual property rights, which are retained by the thesis author, unless otherwise stated.
- A copy can be downloaded for personal non-commercial research or study, without prior permission or charge.
- This thesis cannot be reproduced or quoted extensively from without first obtaining permission in writing from the author.
- The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the author.
- When referring to this work, full bibliographic details including the author, title, awarding institution and date of the thesis must be given.
Evolutionary history, cross-species transmission and host adaptation of human viruses and their primate homologues

Sinéad Lyons

A thesis submitted for the degree of Doctor of Philosophy

The University of Edinburgh

September 2014
Abstract

At present the origins of major human pathogens associated with hepatic disease are poorly understood. The absence of such information pertaining to the evolutionary history of hepatitis B virus (HBV) and hepatitis C virus (HCV) and its genetically related viruses impacts upon the development of vaccines and effective eradication strategies. Studies are currently limited by the absence of historical samples from which to date the emergence of human infections and therefore the evolution of human hepatic viruses relies on epidemiological studies and genetic analysis of contemporary virus populations worldwide.

Approximately one third of the world’s population is infected with HBV, and despite the availability of a vaccine, the virus is attributed with over 1 million deaths per year through liver disease. HBV variants infecting humans show genetic and antigenic heterogeneity and are currently classified into 8 genotypes A-H with nucleotide divergence of between 9-13%. In addition to these variants, recombination has been detected between genotypes A and D, and B and C, which can generate novel variants. The past 10 years has seen the detection of HBV in chimpanzees, gorillas and other non-human primates (NHPs) at frequencies comparable to those observed in regions of endemic human HBV infection. Despite the genetic divergence between human and NHP HBV variants the detection of recombination between human genotype C and chimpanzee and gibbon variants suggests that HBV can share hosts in nature. The evolutionary process that may have given rise to the distinct species-specific variants of NHP HBV within overlapping geographical regions has not been reconciled, with evidence supporting both allopatric speciation and co-speciation.

HCV a member of the Flaviviridae family currently infects approximately 3% of the world’s population and is one of the major causes of chronic liver disease, hepatocellular carcinoma and liver cirrhosis. Human pegivirus (HPgV) a member of the Pegivirus genus of the Flaviviridae family infects approximately 5% of the world’s population, although it
is of unknown disease association. Very recently, several studies of wild rodent and bat populations have revealed much greater viral diversity of members of both Hepacivirus and Pegivirus genera. Homologues of HCV have been detected in a range of species including domestic dogs (canine hepacivirus [CHV]) and horses (non-primate hepaciviruses [NPHV]). Similarly, several new pegiviruses have been described in horses (equine pegivirus, [EPgV] and Theiler’s Disease Associated Virus [TDAV]), several species of rodents (rodent pegivirus [RPgV]), and further species of bats (bat pegivirus, [BPgV]). Despite the differences in pathogenicity between HCV and HPgV infections, they share similar genomic organisation and are capable of establishing persistent infections in humans. Studies into bat, horse and rodent homologs of HCV and HPgV have yet to determine disease associations, transmission routes and seroprevalence.

Studies presented within this thesis broaden our understanding of the clinical presentations and host range of NPHV and EPgV. Screening to determine the level of active and past infection to both viruses provides novel insight into infection frequencies, host range, disease progression and examines the correlation between infections and the presence or absence of hepatic disease. Research examining HBV variants circulating in NHPs in Cameroon provides novel evidence for the occurrence of recombination and cross species transmission between NHP variants of HBV and examines the role these findings play in expanding our understanding of the evolution of HBV.
Declaration of Originality

I declare that the work recorded in this thesis was completed entirely by myself at the Centre for Infectious Diseases and Roslin Institute of the University of Edinburgh between September 2010 and March 2013. In cases where others have made a contribution to the results obtained, this is clearly stated within the text. This work has not been submitted for any other degree or professional qualification. In cases where portions of the work contained within this thesis have been published in academic journals, this is indicated within the text and copies of papers are given as appendices. Where required by license agreement, permission to include published papers has been sought from publishers.

Sinéad Lyons
Acknowledgements

These pages are my small dedication to the countless number of family, friends, colleagues and mentors, who have inspired me, taught me, laughed with me, and offered me their immense support and love.

For giving me the opportunity to come to this wonderful group in the beautiful city of Edinburgh my thanks go to my supervisor Professor Peter Simmonds. You have over the past four years extended to me your unending patience, knowledge, enthusiasm and dedication. I will be forever grateful for the chance you gave to a biochemist to broaden her scientific horizons, abilities, and her future.

I want to extend a special thank you to those who have helped me over the past four years in developing my academically and as a scientist. To Dr Colin Sharp who has always been supportive and full of advice, and who in the first few months provided me with invaluable skills and training that have taken me through my PhD and I will carry into the future. I extend special thanks to Dr Bill Gregory who in the past year has been a wonderful friend, and a great bouncing board for all things protein related. Jeroen, you are a fountain of knowledge, an inspiration in the lab and it has been a pleasure to troubleshoot western blots and SDS Page gels with you. To all members of the group, past, present and returning, Chloe, Djung, Elly, Inga, Nora, Shoko, Nicky, Nigel, Jill, Donald and Carol, thank you. During our time in Summerhall, Kings Buildings and Roslin each one of you has in your special way generously shared your wisdom and experience, offered unending support, and has always been prepared to lift me up on a cloudy day, for which I am will always be grateful. To all the students I have had the pleasure to meet and share what I have learned with, I wish you the best of luck down the paths you choose to travel and thank you for your smiles, your enthusiasm, and for broadening my teaching experience.

For my parents, I do not even know where to start to express how much you have given me. You both have been an inspiration in every day of my life, and if I can be half the person you are I will be truly lucky. I have been blessed over the past few years by your unfloundering support, your friendship, your passion and curiosity, your belief in me, and most of all for that unending love which carries me through every moment of life. Without you none of this would have been possible and you have given me every opportunity in life to find my path.
and have walked with me through rough and smooth. Thank you for always making me feel
so special. I cannot forget my younger brother, who without knowing it has inspired me; I
have always looked up to you and I thank you sincerely for the knowledge that you are
always looking after my back. To all my family and friends, thank you, you will never
understand how much your love, understanding and support means to me, especially Norah,
Terry, uncle Vinny and dear Sophia. To my granddad in the stars who encouraged me to
venture further and walk the road less travelled, thank you for always twinkling over me and
lighting up every day, I miss you.
Finally I want to thank someone who is not just my partner, but my best friend, my shoulder,
my companion on photographic trips, my soldier and the one person who shares all my cares,
worries, doubts and sweeps them all away. Rennos, thank you for so patiently listening to
my rants, for tolerating my frustrations, tiredness, singing and peculiarities, for introducing
me to wonderful Greeks, for letting me share your heart of gold and for making Edinburgh
home. In the past few years you have opened up the world to me and made my life complete,
but most importantly you have loved me and have given me the knowledge that I face
nothing alone.

“I shall be telling this with a sigh
Somewhere ages and ages hence:
Two roads diverged in a wood, and I—
I took the one less travelled by,
And that has made all the difference.”
    Robert Frost
List of Original Publications

Original publications directly associated with this thesis


Original publications to which the author has otherwise contributed


<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>AdV</td>
<td>adenovirus</td>
</tr>
<tr>
<td>A&amp;E</td>
<td>Accident and Emergency department</td>
</tr>
<tr>
<td>AIDS</td>
<td>Acquired Immune Deficiency Syndrome</td>
</tr>
<tr>
<td>ALT</td>
<td>Alanine Aminotransferase</td>
</tr>
<tr>
<td>ASP</td>
<td>Aspartate Aminotransferase</td>
</tr>
<tr>
<td>AP</td>
<td>Antarctic Phosphatase</td>
</tr>
<tr>
<td>APC</td>
<td>Antigen Presenting Cell</td>
</tr>
<tr>
<td>ApoE</td>
<td>Apolipoprotein E</td>
</tr>
<tr>
<td>ARFP</td>
<td>Alternate Reading Frame Protein</td>
</tr>
<tr>
<td>Arg</td>
<td>Arginine</td>
</tr>
<tr>
<td>ART</td>
<td>Antiretroviral therapy</td>
</tr>
<tr>
<td>BHV</td>
<td>Bat hepacivirus</td>
</tr>
<tr>
<td>BLAST</td>
<td>Basic Local Alignment Search Tool</td>
</tr>
<tr>
<td>Bp</td>
<td>Base pair</td>
</tr>
<tr>
<td>BPgV</td>
<td>Bat pegivirus</td>
</tr>
<tr>
<td>BSA</td>
<td>Bovine Serum Albumin</td>
</tr>
<tr>
<td>C</td>
<td>Cytosine</td>
</tr>
<tr>
<td>CAMV</td>
<td>Carnation mottle virus</td>
</tr>
<tr>
<td>CARD</td>
<td>Caspase Recruitment Domain</td>
</tr>
<tr>
<td>cDNA</td>
<td>complementary deoxyribonucleic acid</td>
</tr>
<tr>
<td>cccDNA</td>
<td>covalently closed circular DNA</td>
</tr>
<tr>
<td>CHV</td>
<td>Canine hepacivirus</td>
</tr>
<tr>
<td>CREs</td>
<td>cis-acting replication element</td>
</tr>
<tr>
<td>Cys</td>
<td>cysteine</td>
</tr>
<tr>
<td>CT</td>
<td>cycle threshold for real-time PCR</td>
</tr>
<tr>
<td>DDBJ</td>
<td>DNA Databank of Japan</td>
</tr>
<tr>
<td>DNA</td>
<td>deoxyribonucleic acid</td>
</tr>
<tr>
<td>dNTP</td>
<td>deoxynucleotide triphosphates</td>
</tr>
<tr>
<td>DR</td>
<td>Direct repeat</td>
</tr>
<tr>
<td>dsDNA</td>
<td>double-stranded DNA</td>
</tr>
<tr>
<td>dsRNA</td>
<td>double stranded ribonucleic acid</td>
</tr>
<tr>
<td>DTT</td>
<td>Dithiothreitol</td>
</tr>
<tr>
<td>EB</td>
<td>Elution Buffer</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>-------------</td>
</tr>
<tr>
<td><strong>E. coli</strong></td>
<td><em>Escherichia coli</em></td>
</tr>
<tr>
<td>EDTA</td>
<td>ethylenediaminetetraacetic acid</td>
</tr>
<tr>
<td>eIF</td>
<td>eukaryotic Initiation Factor</td>
</tr>
<tr>
<td>ELISPOT</td>
<td>Enzyme-Linked Immunospot</td>
</tr>
<tr>
<td>EMBL</td>
<td>European Molecular Biology Laboratory</td>
</tr>
<tr>
<td>EN</td>
<td>Enhancer element</td>
</tr>
<tr>
<td>EPGV</td>
<td>Equine pegivirus</td>
</tr>
<tr>
<td>ER</td>
<td>Endoplasmic Reticulum</td>
</tr>
<tr>
<td>EtBr</td>
<td>Ethidium Bromide</td>
</tr>
<tr>
<td>EVR</td>
<td>Early Virological Response</td>
</tr>
<tr>
<td>GARD</td>
<td>genetic algorithm for recombination detection</td>
</tr>
<tr>
<td>GBV-A</td>
<td>GB Virus A</td>
</tr>
<tr>
<td>GBV-B</td>
<td>GB Virus B</td>
</tr>
<tr>
<td>GBV-C</td>
<td>GB Virus C</td>
</tr>
<tr>
<td>γGT</td>
<td>gamma Glutamine Transferase</td>
</tr>
<tr>
<td>GLDH</td>
<td>Glutamate Dehydrogenase</td>
</tr>
<tr>
<td>GST</td>
<td>Glutathione S-Transferase</td>
</tr>
<tr>
<td>GTP</td>
<td>Guanosine Triphosphate</td>
</tr>
<tr>
<td>HAV</td>
<td>Hepatitis A Virus</td>
</tr>
<tr>
<td>HBV</td>
<td>Hepatitis B Virus</td>
</tr>
<tr>
<td>HCV</td>
<td>Hepatitis C Virus</td>
</tr>
<tr>
<td>HCVcc</td>
<td>cell culture produced HCV</td>
</tr>
<tr>
<td>HCVpp</td>
<td>HCV pseudoparticles</td>
</tr>
<tr>
<td>H2O</td>
<td>Water</td>
</tr>
<tr>
<td>HIV</td>
<td>Human Immunodeficiency Virus</td>
</tr>
<tr>
<td>HKY</td>
<td>Hasegawa-Kishino-Yano model of nucleotide substitution</td>
</tr>
<tr>
<td>HLA</td>
<td>Human Leukocyte Antigen</td>
</tr>
<tr>
<td>HPA</td>
<td>Health Protection Agency</td>
</tr>
<tr>
<td>HPgV</td>
<td>Human pegivirus</td>
</tr>
<tr>
<td>HMPV</td>
<td>human metapneumovirus</td>
</tr>
<tr>
<td>HVR</td>
<td>Hyper Variable Region</td>
</tr>
<tr>
<td>ICTV</td>
<td>International Committee on the Taxonomy of Viruses</td>
</tr>
<tr>
<td>IFA</td>
<td>Immunofluorescent Assay</td>
</tr>
<tr>
<td>indels</td>
<td>insertions and deletions in nucleotide sequences</td>
</tr>
<tr>
<td>IDU</td>
<td>Injecting Drug User</td>
</tr>
<tr>
<td>ISDR</td>
<td>Interferon Sensitivity-Determining Region</td>
</tr>
<tr>
<td>IFN</td>
<td>Interferon</td>
</tr>
<tr>
<td>IFNAR I</td>
<td>FN-α/β Receptors</td>
</tr>
</tbody>
</table>
Abbreviations

IR  Insulin Resistance
IRES  Internal Ribosome Entry Site
IRF  IFN Regulatory Factor
IRES  Internal Ribosome-Entry Site
ISG  IFN Stimulated Gene
ISGF3  IFN Stimulated Gene Factor 3
Kb  kilobase
L  leader protein
L  Litre
LD  Lipid Droplet
LDL  low density lipoprotein
MAVS  Mitochondrial Antiviral Signaling Protein
MCL  maximum composite likelihood
MgCl₂  Magnesium Chloride
MgSO₄  Magnesium Sulfate
MHC  Major Histocompatibility Complex
M  Millilitre
MSM  Men who have Sex with Men
MCMC  Markov Chain Monte Carlo
MEGA  Molecular Evolutionary Genetics Analysis
nAb  neutralizing Antibody
NANBH  Non-A, non-B Viral Hepatitis
NAT  Nucleic acid testing
NCBI  National Center for Biotechnology Information
NES  Nuclear Export Signal
NF-κB  Nuclear Factor κB
NHL  Non-Hodgkin Lymphoma
NHP  Non-human primate
NJ  neighbour joining
NNI  Non-Nucleoside Inhibitor
NPHV  Nonprimate hepacivirus
NS  Non-structural
nt  nucleotide
NTPase  RNA nucleoside triphosphatases
OD  Optical Density
ORF  Open Reading Frame
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>OTU</td>
<td>operational taxonomic unit</td>
</tr>
<tr>
<td>PAMPs</td>
<td>Pathogen-Associated Molecular Patterns</td>
</tr>
<tr>
<td>PBS</td>
<td>Phosphate Buffered Saline</td>
</tr>
<tr>
<td>PCR</td>
<td>Polymerase Chain Reaction</td>
</tr>
<tr>
<td>pgRNA</td>
<td>pre-genomic RNA</td>
</tr>
<tr>
<td>PI</td>
<td>Protease Inhibitor</td>
</tr>
<tr>
<td>PKR</td>
<td>dsRNA-activated Protein Kinase R</td>
</tr>
<tr>
<td>PPT</td>
<td>Polypyrimidine Tract</td>
</tr>
<tr>
<td>rcDNA</td>
<td>relaxed circular DNA</td>
</tr>
<tr>
<td>RDP</td>
<td>Recombination detection program</td>
</tr>
<tr>
<td>RdRp</td>
<td>RNA-dependent RNA polymerase</td>
</tr>
<tr>
<td>REMs</td>
<td>Replication Enhancing Mutations</td>
</tr>
<tr>
<td>RHV</td>
<td>Rodent hepacivirus</td>
</tr>
<tr>
<td>RIG-I</td>
<td>Retinoic Acid Inducible Gene-I</td>
</tr>
<tr>
<td>RNA</td>
<td>Ribonucleic Acid</td>
</tr>
<tr>
<td>RpgV</td>
<td>Rodent pegivirus</td>
</tr>
<tr>
<td>rpm</td>
<td>rounds per minute</td>
</tr>
<tr>
<td>RT</td>
<td>Reverse-Transcription</td>
</tr>
<tr>
<td>RVR</td>
<td>Rapid Virological Response</td>
</tr>
<tr>
<td>RT-PCR</td>
<td>reverse transcription polymerase chain reaction</td>
</tr>
<tr>
<td>SC</td>
<td>SuperScriptTM III One-Step RT-PCR</td>
</tr>
<tr>
<td>SCID</td>
<td>Severe Combined Immunodeficiency</td>
</tr>
<tr>
<td>SDS-PAGE</td>
<td>Sodium Dodecyl Sulphate Polyacrylamide Gel Electrophoresis</td>
</tr>
<tr>
<td>shRNA</td>
<td>short hairpin RNA</td>
</tr>
<tr>
<td>siRNA</td>
<td>small interfering RNA</td>
</tr>
<tr>
<td>SN</td>
<td>Supernatant</td>
</tr>
<tr>
<td>SR-BI</td>
<td>Scavenger Receptor Class B Type I</td>
</tr>
<tr>
<td>ssRNA</td>
<td>single-stranded RNA</td>
</tr>
<tr>
<td>STAT</td>
<td>Signal Transducers and Activators of Transduction</td>
</tr>
<tr>
<td>STAT-C</td>
<td>Specifically Targeted Antiviral Therapy for Hepatitis C</td>
</tr>
<tr>
<td>SVR</td>
<td>Sustained Virological Response</td>
</tr>
<tr>
<td>SBP</td>
<td>Single breakpoint recombination detection</td>
</tr>
<tr>
<td>T</td>
<td>Thymine</td>
</tr>
<tr>
<td>TAE</td>
<td>tris-acetate EDTA</td>
</tr>
<tr>
<td>TAH</td>
<td>Transfusion-Associated Hepatitis</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>------------------------------------</td>
</tr>
<tr>
<td>TLR</td>
<td>Toll-Like Receptor</td>
</tr>
<tr>
<td>TMD</td>
<td>Transmembrane Domain</td>
</tr>
<tr>
<td>Tm</td>
<td>melting temperature</td>
</tr>
<tr>
<td>UTR</td>
<td>Untranslated Region</td>
</tr>
<tr>
<td>VLDL</td>
<td>Very Low Density Lipoprotein</td>
</tr>
<tr>
<td>VPU</td>
<td>Viral Protein U</td>
</tr>
<tr>
<td>U</td>
<td>Units</td>
</tr>
<tr>
<td>U</td>
<td>Uracil</td>
</tr>
<tr>
<td>UV</td>
<td>Ultra-Violet</td>
</tr>
<tr>
<td>VP</td>
<td>Viral protein</td>
</tr>
<tr>
<td>VPg</td>
<td>Viral Protein genome linked</td>
</tr>
<tr>
<td>VZV</td>
<td>varicella zoster virus</td>
</tr>
<tr>
<td>WB</td>
<td>Western blot</td>
</tr>
<tr>
<td>WHO</td>
<td>World Health Organisation</td>
</tr>
<tr>
<td>WT</td>
<td>Wild-Type</td>
</tr>
</tbody>
</table>
List of Figures

FIGURE 1.1 Bayesian phylogenetic analysis of the hepadnaviridae family, including human HBV and isolates from other species including rodents, birds, non-human primates, and bats

FIGURE 1.2 The structure of the Hepatitis B Virus virion.

FIGURE 1.3 The genome organisation of the Hepatitis B Virus.

FIGURE 1.4 Current genera of the Flavirviridae family and their respective members.

FIGURE 1.5 Genome organisations of members of the Hepacivirus genus.

FIGURE 1.6 Genome organisations of members of the Pegivirus genus.

FIGURE 2.1 Unidirectional laboratory systems for prevention of PCR contamination in reactions.

FIGURE 2.2 Histopaque 1077 isolation of peripheral blood mononuclear cells from whole blood.

FIGURE 2.3 Vector maps of pGex-2t and pET28B plasmids used in the generation of NPHV and EPgV recombinant proteins.

FIGURE 3.1 Generation of full-length HBV genomes in this study.

FIGURE 3.2 Phylogenetic analyses based on the complete HBV genome.

FIGURE 3.3 Mitochondrial sequencing of HBV isolates in this study to determine species and sub-species of NHP. Reference Gorilla gorilla, P.t. elliotti and P.t. troglodytes were included in SSE.

FIGURE 3.4 Tree Order Scan of NHP HBV sequences with reference sequences from all human HBV genotypes.

FIGURE 3.5 Tree Order Scan analysis studies HBV isolates with available NHP HBV sequences.
FIGURE 3.6 Phylogenetic analyses based on the recombinant region 1560-2120bp of HBV genome

FIGURE 3.7 Grouping Scan analysis of (A) Gorilla HBV study recombinant compared with (B) non-recombinant chimpanzee AM117396

FIGURE 3.8 Grouping Scan analyses of HBV recombinants from P. t. elliotti and P. t. troglodytes identified in this study.

FIGURE 3.9 Grouping scan analysis of HBV recombinants AB498266 and AB046525 from P. t. schweinfurthii and P. t. troglodytes respectively.

FIGURE 3.10 Geographical locations of suggested wild capture of study isolates in Cameroon.

FIGURE 3.11 Geographical distribution and current distribution of Pan. in East Africa

FIGURE 4.1 Phylogenetic analysis of A) 5’ untranslated region, B) nonstructural protein 3, and C) nonstructural protein 5B regions of nonprimate hepatitis virus sequence amplified from screen-positive study animals. Neighbour-joining trees of nucleotide sequences from each genome region

FIGURE 4.2 Phylogenetic analyses of NPHV isolates in (A) NS3 and (B) NS5B region in an extended dataset with representative sequences from novel BHV, RHV and GHV variants, and HCV genotypes 1-7

FIGURE 5.1 Phylogenetic analyses of EPgV study isolates in the NS3 helicase region with representative members of the Pegivirus genus.

FIGURE 5.2 Phylogenetic analysis of the NS5B region of EPgV study isolates with representative members of the Pegivirus genus

FIGURE 5.3 StructureDist Dot Plot predicting secondary structural elements in the 5’UTR of EPgV.

FIGURE 5.4 Consensus model for the secondary structure of EPgV 5’UTR using combined PFOLD, ALIFOLD and Structure Dist programs
FIGURE 6.1  SDS-PAGE analysis of unpurified *E. coli* lysate with recombinant proteins generated to detect NPHV and EPgV antibodies.

FIGURE 6.2  Western blot detection of EPgV Anti-NS3 antibodies in ELISA seropositive animals.

FIGURE 6.3  Western blot analysis of unpurified *E. coli* soluble lysate with recombinant proteins generated to detect NPHV-NS3/Core and EPgV-NS3 antibodies.

FIGURE 6.4  Western blot detection of EPgV Anti-NS3 antibodies in ELISA seropositive animals.

FIGURE 6.5  Demonstrated seroreactivity of serum tested in non-expressing control wells versus (A) NPHV-helicase, (B) NPHV-core and (C) EPgV-helicase antigen coated wells.

FIGURE 6.6  Anti-NPHV helicase and core IgG detection across range of species.

FIGURE 6.7  Assay validations of NPHV ELISA and correlation between samples identified as NS3 and/or Core antibody positive.

FIGURE 6.8  Anti-NPHV helicase and core IgG detection in HCV positive and HCV negative human serum.

FIGURE 6.9  Anti-EPgV helicase IgG detected in horses by ELISA.
**List of Tables**

| TABLE 1.1  | Estimated length of the 5’UTR and 3’UTR of members of the *Hepacivirus* genus |
| TABLE 1.2  | Published estimates of the 5’UTR and 3’UTR of identified members of the *Pegivirus* genus |
| TABLE 1.3  | Generalised functions of the hepacivirus and pegivirus encoded proteins in the viral life cycle |
| TABLE 1.4  | Current nomenclature, classification and pathogenesis of original ‘GB’ viruses and hepaciviruses |
| TABLE 1.5  | Aetiological agents of equine hepatic insufficiencies, their sources and the clinical indicators resulting |
| TABLE 2.1  | Reagents included in reverse transcription reactions |
| TABLE 2.2  | Superscript III RT-PCR Reagents |
| TABLE 2.3  | SuperScript III Protocol |
| TABLE 2.4  | Codes for individual nucleotides and degenerate bases used for PCR primer design |
| TABLE 2.5  | GoTaq PCR reagents |
| TABLE 2.6  | GoTaq PCR protocol |
| TABLE 2.7  | AccuPrime Reagents |
| TABLE 2.8  | AccuPrime PCR Protocol |
| TABLE 2.9  | Hot Star HiFi PCR Reagents |
| TABLE 2.10 | Hot Star HiFi PCR Protocol |
List of Tables

TABLE 2.11  Hot Star PCR Protocol
TABLE 2.12  Hot Star PCR Reagents
TABLE 2.13  Modifications to standard PCR protocols used for individual PCR reactions
TABLE 2.14  Sequencing reactions reagents
TABLE 2.15  Sequencing PCR conditions
TABLE 2.16  Reaction components of pGem T-Easy vector system incubated overnight at 4°C
TABLE 2.17  Reaction components of MEGAscript T7 Kit for generation of RNA transcripts
TABLE 2.18  Restriction Digest Reagents
TABLE 2.19  Restriction Digest Conditions
TABLE 2.20  Modifications to Restriction Digest Reagents and Conditions
TABLE 2.21  Ligation reaction reagents used in the construction of recombinant plasmids containing NPHV and EPgV viral protein coding sequences
TABLE 3.1  Old World and New World primates and identified HBV variants
TABLE 3.2  HBV primers designed in this study to generate complete genomes
TABLE 3.3: Mitochondrial specific primers to determine NHP species and sub-species
TABLE 3.4  Genbank Accession Numbers of human HBV genotypes analysed in phylogenetic analysis with NHP HBV isolates

TABLE 3.5  Specimen isolation data for all HBV positive NHPs

TABLE 4.1  Titration of NPHV-NS3 transcripts tested with NS3 primers to determine PCR sensitivity for NPHV RNA in samples from domestic horses in the United Kingdom

TABLE 4.2  Nonprimate hepacivirus primer sequences for 5’UTR, NS3, and NS5B in samples from horses, United Kingdom

TABLE 4.3  Nonprimate hepacivirus sequences detected by using PCR on mammal samples, United Kingdom

TABLE 4.4  Clinical features of domestic horses infected by nonprimate hepaciviruses, United Kingdom

TABLE 4.5  Laboratory indices for domestic horses infected with nonprimate hepaciviruses, United Kingdom

TABLE 5.1  Transcript titration of NPVH and EPgV NS3 helicase controls

TABLE 5.2  Detection of NPHV, EPgV and TDAV RNA in non-human primates

TABLE 5.3  Sample information for all horses testing RNA positive for active EPgV infection.

TABLE 5.4  Nucleotide and amino acid (shaded) divergence in the NS3 and NS5B regions between EPgV isolates and known sequences from each of the species-specific clades; RPgV, BPgV, SPgV and HPgV

TABLE 5.5  Viral load analysis and liver enzyme data for EPgV RNA positive horses.

TABLE 5.6  Identified sites of covariance and semi-variance, insertions, deletions, disruptive nucleotide pairings and unpaired sites in the preliminary modelling of the 5’UTR of EPgV.

TABLE 6.1  Details of recombinant plasmids generated, and predicted molecular weights, confirmed by SDS-PAGE analysis
**List of Tables**

TABLE 6.2  NPHV RNA and antibodies detected by PCR and ELISA respectively

TABLE 6.3  Seropositive horses determined by detection above cut-off of antibodies to both NS3 and Core NPHV proteins by ELISA.

TABLE 6.4  Relative numbers of horses testing antibody positive to NS3 and Core ELISAs and/or RNA positive for NPHV.

TABLE 6.5  NPHV NS3 and core antibodies detected in five horses and one dog cohabiting with an RNA positive horse.

TABLE 6.6  Detection of TDAV and EPgV RNA and antibodies by virus specific PCR and EPgV NS3 ELISA respectively.

TABLE 6.7  Relative numbers of EPgV seropositive horse testing antibody positive by western blot

TABLE 6.8  The relative number of horses testing antibody positive and/or RNA positive for EPgV.

TABLE 6.9  Detection of EPgV NS3 antibodies by ELISA and confirmed by western blot in RNA positive horses.

TABLE 6.10  The relative number of horses confirmed as antibody positive for both EPgV and NPHV.
# Contents

Abstract .................................................................................................................. i  
Declaration of Originality ....................................................................................... iii  
Acknowledgements ................................................................................................. iv  
List of Original Publications .................................................................................... vi  
Abbreviations .......................................................................................................... vii  
List of Figures .......................................................................................................... xii  
List of Tables ............................................................................................................ xv  
Contents ................................................................................................................... xix  

**Chapter 1 Introduction** ...................................................................................... 1  
1.1 Molecular virology of Hepatitis B Virus ............................................................... 1  
1.1.1 **Hepadnaviridae** family ................................................................................. 1  
1.1.2 HBV Virion structure and genome organisation ............................................ 2  
1.1.3 Functions of encoded proteins .................................................................... 5  
1.1.4 HBV Replication .......................................................................................... 9  
1.1.5 Genetic Diversity of HBV ............................................................................ 10  
1.1.6 HBV in NHP primates ................................................................................. 15  
1.1.7 Recombination within and between HBV variants infecting humans and non-human primates ................................................................................. 17  
1.1.8 Proposed models and hypotheses of HBV evolution .................................... 19  
1.1.9 Mechanisms of HBV recombination ........................................................... 22  
1.2 Hepatic viruses of the **Flaviviridae** family ...................................................... 24  
1.2.1 **Flaviviridae** family and classification ..................................................... 24  
1.2.2 Genome organisation of **Hepacivirus** and **Pegivirus** genus ................. 27  
Hepacivirus-HCV Genome ...................................................................................... 27  
1.2.3 Proteolytic cleavage of HCV and HPgV and related viruses ...................... 32  
1.2.4 Functions of encoded viral proteins ............................................................ 33  
1.2.5 Translation .................................................................................................. 39  
1.2.6 RNA replication .......................................................................................... 39  
1.2.7 Transmission routes associated with hepacivirus and pegivirus infection ..... 40  
1.3 History and Origins of Hepatitis C Virus ......................................................... 43  
1.3.1 Genetic diversity and genotype classification ............................................. 44  
1.3.2 Dissemination of HCV worldwide .............................................................. 45  
1.3.3 HCV geographical distribution ................................................................. 46  
1.3.4 Homologs of HCV in other mammals ....................................................... 48  
1.3.5 Disease association of hepacivirus infections ............................................ 50  
Chronic hepatitis .................................................................................................... 52  
1.3.6 Prevalence of hepaciviruses in mammalian species .................................... 53  
1.4 History of human pegivirus .............................................................................. 55
Chapter 2 Materials and Methods

1.5 Other points of consideration .......................................................... 63
1.5.1 Diagnostic methods applied in detection of equine hepatopathy ........ 63
1.5.2 Causative agents associated with equine hepatic disease .................. 66
1.6 Aims of this thesis ........................................................................... 69

Chapter 3 Species Association and HBV recombination between Non-human primate variants

3.1 Introduction ..................................................................................... 113
3.2 Materials and Methods .................................................................. 117
3.2.1 Sample selection ........................................................................ 117
3.2.2 HBV screening and whole genome sequencing ......................... 118
3.2.3 Determination of chimpanzee sub-species ................................. 120
3.2.4 Sequencing of PCR products ...................................................... 121
3.2.5 Phylogenetic Analysis ................................................................. 121
3.2.6 Recombination Analysis .............................................................. 122
3.3 Results ............................................................................................ 123
3.3.1 Isolation of HBV from wild caught non-human primates samples between August 2004 and August 2009, Cameroon .................................................. 123
3.3.2 Determination by phylogenetic analysis of relationship between NHP HBV isolates .................................................................................................................. 125
3.3.3 Mitochondrial sequencing for determination of NHP species and sub-species 127
3.3.4 Recombination detection by Tree order scan analysis across HBV genome 128
3.3.5 Determination of recombination breakpoints in HBV genome of NHP study isolates .................................................................................................................. 134
3.4 Discussion .................................................................................................................. 138
3.4.1 Screening of NHPs for presence of human and species-specific variants.... 138
3.4.2 Complete genome analysis of study sequences isolates from NHPs........ 138
3.4.3 Detection of HBV recombination between _Gorilla gorilla/ P. t. troglodytes_. 141
3.4.4 Co-divergence and allopatric speciation as models of HBV evolution....... 142
3.4.5 Emergence of potential reservoir for primate hepadnaviruses ................. 145
3.4.6 Conclusions ............................................................................................................ 146

Chapter 4 Non-primate hepacivirus; detection and characterisation in domestic horses in the United Kingdom.................................................................................................................. 148
4.1 Introduction................................................................................................................. 148
4.1.1 Pandemic spread of HCV in the Western world............................................. 148
4.1.2 HCV genetic diversity and endemic circulation............................................ 149
4.1.3 Characterisation of HCV homologues in other mammals........................... 150
4.2 Materials and Methods............................................................................................ 151
4.2.1 Selection of samples ....................................................................................... 151
4.2.2 CHV and NPHV Screening ............................................................................ 152
4.2.3 Sequencing of PCR amplicons ...................................................................... 155
4.2.4 Determination of NPHV viral load................................................................. 155
4.2.5 Sequence alignment and phylogenetic analysis............................................. 156
4.2.6 Clinical history and liver enzyme data ......................................................... 156
4.3 Results....................................................................................................................... 156
4.3.1 Screening for NPHV and homologs in range of mammalian species ......... 156
4.3.2 Phylogenetic analysis of NPHV variants with known members of the _Hepacivirus_ genus ........................................................................................................ 158
4.3.3 Clinical characteristics and disease progression of NPHV-positive Horses . 162
4.4 Discussion ................................................................................................................. 164
4.4.1 Detection of Non-primate hepacivirus in UK Horses .................................. 164
4.4.2 Genetic relationship between NPHV and other members of the _Hepacivirus_ genus............................................................................................................. 166
4.4.3 Clinical presentations of NPHV infections ................................................... 167
4.4.4 Evidence for establishment of persistent infection...................................... 168
4.4.5 Conclusions ........................................................................................................ 170

Chapter 5 Epidemiology, Clinical Presentations and structural analysis of 5'UTR of EPGV................................................................................................................................. 172
5.1 Introduction ........................................................................................................................................... 172
  5.1.1 Host range and genetic diversity of members of the *Pegivirus* genus ........................................... 172
  5.1.2 Frequency and distribution of pegivirus infection in humans and NHPs ......................................... 173
  5.1.3 Clinical indications of pegivirus infection ...................................................................................... 173
  5.1.4 Progression of pegivirus infection in the horses ............................................................................ 175
  5.1.5 *Pegivirus* 5’UTR: Secondary structure and type 4 internal ribosomal entry site (IRES) ............... 175
5.2 Materials and Methods ........................................................................................................................ 176
  5.2.1 Sample selection .......................................................................................................................... 176
  5.2.2 NPHV, EPgV, and TDAV Screening ............................................................................................ 177
  TDAV PCR, RNA transcripts were generated from a plasmid containing PCR amplified partial NS3 cDNA by using the Ambion T7 transcription kit (Promega Corp., Southampton, UK). Transcripts were purified with the RNeasy kit (QIAGEN), and concentrations were determined by using the NanoDrop 2000 (NanoDrop Products, Wilmington, DE, USA). RNA extractions were performed on 140 µl of plasma using the QIAmp viral extraction kit (QIAGEN) according to the manufacturer's instructions and eluted in a final volume of 60 µl ............................................................................ 178
  5.2.3 Viral load quantification ............................................................................................................. 179
  5.2.4 Clinical history and liver enzyme data ......................................................................................... 179
  5.2.5 Determination of 5’UTR secondary structure ............................................................................ 180
5.3 Results ................................................................................................................................................ 180
  5.3.1 Hepacivirus infections in non-human primates ............................................................................ 180
  ND; not done ........................................................................................................................................ 181
  5.3.2 Pegivirus infections in horses .................................................................................................... 181
  5.3.3 Phylogenetic analysis of novel EPgV sequences in conserved NS3 and NS5B regions of the pegivirus genome ................................................................................................................................. 182
  5.3.4 Clinical characterisation of EPgV infection ................................................................................. 187
  5.3.5 Frequency of TDAV infections ................................................................................................... 190
  5.3.6 Secondary structure of EPgV 5’UTR .......................................................................................... 190
5.4 Discussion .......................................................................................................................................... 195
  5.4.1 Active hepacivirus infection in domestic horses in the UK ......................................................... 195
  5.4.2 Frequency of pegivirus infections in the domestic horse population .......................................... 195
  5.4.3 Phylogenetic analysis determines species specific clade of EPgV ............................................. 197
  5.4.4 Analysis of correlation between NPHV and EPgV viraemia and hepatopathy in horses .......... 198
  5.4.5 Absence of hepaci-like viruses in NHPs ...................................................................................... 200
  5.4.6 Secondary structure prediction of 5’UTR and IRES elements of EPgV ..................................... 201
  5.4.7 Conclusion ................................................................................................................................... 202

Chapter 6 Prevalence of NPHV and EPgV antibodies across species ........................................ 204

6.1 Introduction ..................................................................................................................................... 204
6.1.1 Diagnostic methods for the detection of hepacivirus and pegivirus exposure in mammals ......................................................... 204
6.1.2 Hepacivirus transmission between species ........................................ 206
6.1.3 Level of exposure in global population to hepacivirus and pegivirus infections as a measure of antibody reactivity .......................... 207
6.1.4 Antibody detection and the establishment of persistent hepacivirus infections .......................................................... 208
6.2 Materials and Methods .................................................................. 210
6.2.1 Sample selection ........................................................................ 210
6.2.2 Serology ..................................................................................... 211
6.2.3 Enzyme linked immunosorbent assay ........................................... 217
6.2.4 Western blot analysis of EPgV seropositive animals ..................... 218
6.3 Results ........................................................................................... 220
6.3.1 ELISA Screening for past exposure to NPHV across species ............ 220
6.3.2 Preliminary estimation of seroprevalence of EPgV ......................... 227
6.4 Discussion ..................................................................................... 232
6.4.1 Serology based methods for the detection of NPHV and EPgV infection across species .......................................................... 232
6.4.2 Serological evidence to suggest NPHV is endemic in the equine population ............................................................ 234
6.4.3 Evidence for high levels of exposure to and clearance of EPgV ........... 236
6.4.4 Assessment of the correlation between NPHV and EPgV antibody detection and symptoms of hepatic disease ............................ 238
6.4.5 Conclusions .............................................................................. 240

Chapter 7 Concluding remarks ................................................................. 242
7.1 Global importance of elucidating the origins of HBV and HCV ............. 242
7.2 Novel findings of HBV recombination and species association among NHPs .... 244
7.3 Significance of detection and clinical characterisation of NPHV infection .... 246
7.4 Epidemiology and clinical significance of EPgV ................................ 247
7.5 Novel serological assays determine endemic levels of NPHV and EPgV among horses ...................................................................................................................................... 249
7.6 Concluding remarks ........................................................................ 251

References ............................................................................................ 254

Appendix A – List of Primers .................................................................. 308

Appendix B – Supplementary data in the determination of EPgV 5’UTR secondary structure ......................................................................................................................... 311

Appendix C: Permission to reproduce figures ........................................ 315

Appendix D: Transcript Sequences .......................................................... 317

Appendix E: Publications ....................................................................... 318
Chapter 1
Introduction

1.1 Molecular virology of Hepatitis B Virus

1.1.1 Hepadnaviridae family

Hepadnaviruses are a family of double-stranded, enveloped, viruses associated with liver infection in animals and humans. The family consists of two genera, the Orthohepadnavirus and the Avihepadnavirus, of which the hepatitis B virus (HBV) and the duck hepatitis B virus are the type members, respectively. All members of this family contain small partially double-stranded genomes that replicate via an RNA intermediate. Genus Orthohepadnavirus includes HBV isolates from bats, rodents (woodchuck/squirrel), humans and non-human primates (gorillas, chimpanzees [chHBV], woolly monkey [WMHBV]) (Marion et al., 1980, Summers et al., 1978, Grethe et al., 2000b, Norder et al., 1996, Robertson and Margolis, 2002, Sall et al., 2005, Warren et al., 1999, Starkman et al., 2003, Lanford et al., 1998).

Genus Avihepadnavirus includes variants detected in avian species including ducks [DHV], geese, cranes and stork (Guo et al., 2005, Frieman and Cossart, 1986, Robinson et al., 1984).

![Figure 1.1: Bayesian phylogenetic analysis of the hepadnaviridae family, including human HBV and isolates from other species including rodents, birds, non-human primates, and bats. Phylogenetic analysis was based on complete genomes and the space between orthohepadnaviruses](image-url)
Hepatitis B virus, originally termed ‘serum hepatitis’ (Barker et al., 1996, Crossart, 1972), the type member of the Hepadnaviridae family is a small 30-42nm, enveloped, partially double-stranded, relaxed circular DNA virus (rcDNA) with genome of 3.2kb (Figure 1.2). The rcDNA is converted upon infection inside the host cell nucleus, into a plasmid-like covalently closed circular DNA (cccDNA) (Nassal and Schaller, 1996); and subsequently this cccDNA is used to transcribe several genomic and sub-genomic RNAs using host RNA polymerase II; the pregenomic RNA (pgRNA) is selectively packaged into progeny capsids and is reverse transcribed by the virus encoded P protein into new rcDNA genomes.

The virus particle is made up of an outer lipid envelope containing an icosahedral nucleocapsid core, which encloses a single copy of viral DNA and DNA polymerase that is covalently joined to the 5’ end of the minus DNA strand (Figure 1.2). Embedded in the outer lipid membrane are three surface glycoproteins (Large; L, Medium; M and Small; S) involved in viral binding, and entry into susceptible cells. One strand, the ‘minus’ strand ((-) DNA), is almost a complete circle and contains the overlapping genes encoding the structural proteins Pre-S, surface, core and the replicative proteins, polymerase and X. The ‘plus’ strand ((+) DNA) is shorter and can vary in length.
Figure 1.2: The structure of the Hepatitis B Virus virion. The virion consists of an outer lipid envelope and an icosahedral nucleocapsid core. Virions are between 30-42nm in diameter. The nucleocapsid encloses the viral DNA and DNA polymerase.

Each of these viral proteins is translated from one of the four open-reading frames (Figure 1.3). The largest transcript (3.5kb) is the template for genome replication and for the expression of the pre-core/core and polymerase proteins (Summers and Mason, 1982). The pre-S1, pre-S2 and HBV surface antigen is encoded by 2.4kb transcript and the smallest 0.7kb transcript encodes the X protein (Figure 1.3) (Ganem and Schneider, 2001).

The core gene, which contains a pre-core region, encodes for the core antigen HBcAg and the ‘e’ cleavage product and antigen HBeAg. The pre-core/core protein is targeted to the endoplasmic reticulum (ER), and processed at the N- and C-termini, yielding the core protein and releasing the HBeAg. The HBcAg is an
integral component for viral packaging. HBeAg is a good marker for active viral replication as it is generated from the pre-core/core template (Ganem and Schneider, 2001).

The pre-S/surface gene encodes the envelope of the virus and studies have demonstrated its role in receptor binding and initiation of infection (Ganem and Schneider, 2001). Pre-S is required for the attachment of the virus to hepatocytes, however pre-s proteins can also be detected in peripheral blood mononuclear cells (PBMCs) (Zoulim et al.).

Two further regions of the HBV genome contain transcriptional enhancers, EN1 and EN2. EN1 is tissue specific and efficiently replicates only in hepatocytes. The second enhancer element has been shown to be highly liver specific and functions only in highly differentiated human hepatoma cells (Su and Yee, 1992) and stimulates the transcriptional activity of the surface gene promoters. The 5’ strands of both strands of the genome contain short regions of 11 nucleotides, direct repeats, DRs. The 5’ end of the (-) DNA strand maps within the DR1 repeat, and the (+) DNA strand to DR2. Theses repeats are used in priming the synthesis of their respective DNA strands (Chisari and Ferrari, 1997, Beck and Nassal, 2007).
Figure 1.3: The genome organisation of the Hepatitis B Virus. The HBV genome is a partially double stranded DNA virus of approximately 3.2kb. There are four genes encoded by the genome in four overlapping open reading frames (ORFs), C (Core), X, P (Polymerase), and S (Surface).

1.1.3 Functions of encoded proteins

The HBV genome contains 4 ORFs encoding 7 viral proteins, including 3 envelope proteins; S (small), M (medium) and L (large), the replicative P protein (DNA polymerase), the core protein (C) or HBc antigen (HBcAg), the pre-core which is the precursor of the HBe antigen (HBeAg) and the X protein or HBx antigen (HBxAg) (Ocama et al., 2005, Seeger and Mason, 2000)

Envelope

The surface proteins are encoded by a single ORF with three in-frame AUG start codons into the PreS1, PreS2 and S regions. The large ‘L’ protein encompasses all three regions of the ORF (PreS1/S2/S ~389aa), the middle ‘M’ envelope protein
encompasses the PreS2 and S (281aa), while the small ‘S’ protein is encoded by the S domain (226aa) (Ganem and Schneider, 2001).

The three proteins are synthesized at the endoplasmic reticulum (ER) membrane and have a complex trans-membrane topology. All three envelope glycoproteins are anchored to the membrane by the S domain (Eble et al., 1990, Eble et al., 1987). The presence of an un-cleaved N-terminal signal sequence (8-22aa) on the S protein initiates membrane insertion forming the first transmembrane domain (TM1). A second signal sequence at 80-98aa forms TM2 responsible for the translocation of the growing chain through the ER membrane to the ER lumen, until at ~170aa the hydrophobic domain of the C-terminus begins. The C-terminal domain is thought to contain a further 2 TM domains (Eble et al., 1990, Eble et al., 1987).

The M protein can be found in three forms, unglycosylated (p30), monoglycosylated (gp33) and biglycosylated (gp36) (Gerlich et al., 1992, Bruss and Ganem, 1991). In contrast to the S protein the L protein domains PreS1/S2 and part of S remain on the cytosolic side of the ER, and the TM2 domain anchors the growing chain, translocating downstream sequences into the lumen and enabling glycosylation of the S domain (Asn 146) (Bruss et al., 1994, Ostapchuk et al., 1994, Prange and Streeck, 1995). The unused TM1 domain is integrated into the ER and the PreS1-S2 domain faces the ER lumen. Therefore the L protein exists in two topologies: its N-terminal region is either exposed at the surface or oriented to the interior of the virion.

The HBV surface proteins are not just part of the viral particle and research suggests they bud from post-ER pre-Golgi membranes without the envelopment of nucleocapsids into the lumen, and forming subviral particles of 20nm (Huovila et al., 1992). These subviral particles (SVPs) are produced in excess (at least 1000-fold more SVPs than virions) and may interfere with the host immune system and support HBV infection, acting as a decoy and impeding the immune system from effectively
eliminating infectious virions (Bruns et al., 1998).

**Pre-Core/Core**

The 21 kDa core phosphoprotein composes the HBV nucleocapsid, and is conserved between genotypes (Chain and Myers, 2005). The core protein is translated from the pgRNA. The 34 aa of the C-terminal is arginine rich conferring a positive charge and is similar to DNA binding proteins and is associated with the strong interaction between the core protein and the DNA genome, and is therefore essential for packaging of the pgDNA/HBVPol complex (Gallina et al., 1989, Zlotnick et al., 1997). The N-terminal domain is involved in the encapsidation of the viral genome and in its replication. During the viral life cycle the capsid formation is triggered by the binding of the Pol-complexed to the viral pre-genome in conjugation with certain cellular chaperones and kinases (Bartenschlager and Schaller, 1992, Hirsch et al., 1990, Lott et al., 2000, Stahl et al., 2007).

The pre-core protein is produced from a 3.5 kb mRNA (Figure 1.2), producing a primary product of 25kDa. Subsequently this undergoes proteolytic modifications forming the HBeAg, detectable in the serum of infected individuals. The pre-core protein is not necessary for HBV infection and its precise role remains to be determined. At present two hypotheses suggest the pre-core of HBV is implicated in viral persistence. The first hypothesis implies the cytoplasmic fraction of the pre-core protein would diminish the production of mature capsids (Scaglioni et al., 1997), while the second implies that HBeAg induces an immunological tolerance *in utero* (Milich et al., 1990).

**Polymerase**

HBV replicates *via* a pgRNA intermediate of a 3.5kb RNA. The polymerase enzyme associated with replication is the largest of the HBV proteins at 90kDa, containing three domains with four enzymatic activities. The N-terminal domain enables the
priming activity that allows for the initiation of the reverse transcription. A spacer domain between the N-terminal and third domain likely gives the molecule flexibility. The third domain contains the reverse transcriptase, the DNA polymerase and the RNase H activities. The enzyme is encoded by the P ORF and is translated from the pgRNA - which also serves as template for the translation of the core protein – using a leaky scanning mechanism (Fouillot et al., 1993).

HBV Pol has also been implicated in immune evasion, blocking the TLR3- and RIG-I-induced pattern recognition receptor signalling (Wang and Ryu, 2010). Previous studies have documented HBV Pol interaction with host factors such as the Hsp90 facilitating priming, a critical step in viral genome replication (Hu and Seeger, 1996). Further interactions with the Hsp60 chaperone have also been observed and determined to be an important step for maturation of HBV Pol into the active state (Park and Jung, 2001).

X

Discovered over 30 years ago the X protein was so called as no direct function could be assigned or predicted from its amino acid sequence. The X protein is encoded by the X gene and translated from a 0.8kb mRNA and research suggests it is a multifunctional protein. The 154aa protein (17kDa) is conserved among hepadnaviruses infecting mammals. HBx contains a Nuclear Export Signal (NES) motif and a transactivation domain (Forgues et al., 2003, Forgues et al., 2001). HBx is mainly localized in the cytoplasm with a mitochondrial fraction (Henkler et al., 2001). In infected woodchuck hepatocytes small amounts of WHBx (the WHV counterpart of HBx) has been detected in the nuclear matrix (Dandri et al., 1998). HBx has been shown to be a transactivator of cellular and viral genes in vitro, in agreement with its nuclear location. Several groups have proposed that HBx acts as a transcriptional transactivator or interact with transcriptional transactivator(s). Transcriptional activators of the bZip family and Egr1 have been reported to interact
with HBx (Maguire et al., 1991, Williams and Andrisani, 1995, Yoo et al., 1996, Barnabas et al., 1997). Other studies suggest HBx is involved directly or indirectly in the viral infection and/or multiplication, in line with its cytoplasmic localisation. Two studies on WHV have suggested critical role for WHBx in viral replication (Chen et al., 1993, Zoulim et al., 1994, Qadri et al., 1995). The protein has been proposed to have an oncogenic role in HBV infection and the development of hepatocellular carcinoma (Arbuthnot et al., 2000). Through the induction of survivin, disruption of the p53 gene and activation of c-jun N-terminal kinases, HBx suppresses apoptosis of hepatocytes (Marusawa et al., 2003, Kim et al., 2001). All of these proteins play a critical role in the regulation of the cell cycle, and indeed mutations of the p53 gene are evidence of tumorigenesis.

1.1.4 HBV Replication

The replication cycle in HBV is a complex process, as one of only a few classified pararetroviruses, it replicates via a RNA intermediate as observed in retroviruses. HBV is thought to be endocytosed following binding on the cell surface to sodium taurocholate co-transporting polypeptide receptor (NTCP)(Yan et al., 2012). As HBV replication relies on an RNA intermediate produced using host enzymes, the viral genome is transported to the nucleus by host proteins and chaperones following endocytosis. The rcDNA is then made fully double stranded and transformed into cccDNA that serves as the template for the transcription of the four viral mRNAs. The largest of these mRNAs is used to generate new copies of the viral genome in addition to the capsid and DNA polymerase. These four viral transcripts undergo additional processing and go on to form progeny virions that are released from the cell or returned to the nucleus and re-cycled to produce even more copies (Beck and Nassal, 2007, Bruss, 2007). Using cccDNA as a template cellular RNA polymerase II generates pgRNA, longer than the viral genome with an additional 120nt that contain a second copy of the DR1 and ε signals and a poly-A tail. The long mRNA is
then transported back to the cytoplasm where the virion P protein (the DNA polymerase) synthesizes DNA via its reverse transcriptase activity.

Packaging and reverse transcription of pgRNA is modulated by cis-elements of the pgRNA, the encapsidation ε signal and the P protein that binds to ε. This interaction is poorly understood but leads to the recruitment of dimers of the core protein allowing packaging of the pgRNA-P complex. The P-ε interaction is further involved in the initiation of reverse transcription. The first DNA nt becomes covalently linked to the P protein and extended to form the (-) DNA strand and subsequently (+) DNA, producing the rcDNA molecule. Nucleocapsids containing matured rcDNA can be used for intracellular cccDNA amplification, or be enveloped and released from the cell as progeny virions.

1.1.5 Genetic Diversity of HBV

Human HBV is globally distributed and chronically infects approximately 240 million worldwide, one third of the human population and annually between 600,000 and 1 million deaths associated with liver disease or complications of hepatocellular carcinoma are attributed to HBV infection (MacDonald et al., 2000, Ott et al., 2012, Goldstein et al., 2005, Edmunds et al., 1993, McMahon et al., 1985). Currently HBV variants are classified into 8 human genotypes (A-H) (Figure 1.4), with subtypes identified in genotypes A-G and species associated chimpanzee, gorilla and gibbon/orang-utan variants (MacDonald et al., 2000, Hu et al., 2000, Vartanian et al., 2002, Hu et al., 2001, Warren et al., 1999, Grethe et al., 2000b, Verschoor et al., 2001, Noppornpanth et al., 2003, Takahashi et al., 2000). Two putative genotypes I and J have also been reported. Genotype I was tentatively suggested for strains recovered in Laos (Olinger et al., 2008), later shown to fall within the divergence range for intragenotype and not intergenotype divergence (Kramvis et al., 2008). A ninth genotype J was recovered from an 88-year-old Japanese patient with hepatocellular carcinoma, with mean sequence divergence between HBV/J and
gibbon and orangutan genotypes of 10.9% and 10.7% respectively (Tatematsu et al., 2009). HBV variants infecting humans display nucleotide divergence of 9 to 13% between genotypes (Norder et al., 1992a, Norder et al., 1992b, Norder et al., 1993). Okamoto et al analysed 18 full-length genomes and assigned them genotypes A to D (Okamoto et al., 1988). Genotypes were assigned based on sequence divergence of 8% or more. Subsequently four more human genotypes E-H were discovered (Arauz-Ruiz et al., 2002, Naumann et al., 1993, Norder et al., 1994, Stuyver et al., 2000).
Introduction

Figure 1.4: Unrooted phylogenetic tree of 175 full genome sequences of HBV representative of the eight genotypes, using neighbour-joining method. Letters A–H represent the eight genotypes and the geographical restriction of the genotypes is indicated. Phylogenetic analysis indicates distinct genotype specific clades with further geographical associations and sub-genotype associations. Image reproduced with permission from the author (Kramvis et al., 2005).
The 8 HBV genotypes display distinct geographical distributions (Figure 1.5). While genotypes A, D and G are considered to be distributed worldwide, genotype A is predominately found in regions of Northwest Europe, North America and Africa (Norder et al., 1993, Lindh et al., 1997, Lindh et al., 2000) and genotypes B and C are predominately detected in East and South East Asia (Okamoto et al., 1988, Lindh et al., 1997), while genotype D is largely found in regions of the Mediterranean (Westland et al., 2003). Genotype E is found in West Africa and the east coast of Madagascar (Norder et al., 1994, Lindh et al., 1997, Odemuyiwa et al., 2001), and genotypes F is known to be associated with aboriginal populations of South America (Norder et al., 1993, Arauz-Ruiz et al., 1997). Genotype G has been specifically detected among HBV carriers in the UK, USA, Germany, France and Italy (Stuyver et al., 2000, Kato et al., 2002a, Westland et al., 2003, Vieth et al., 2002). Finally genotype H has been found among various indigenous populations in Central and South America including Nicaragua and Mexico and areas of California (Andre, 2000, Arauz-Ruiz et al., 2002, Sanchez et al., 2002) (Figure 1.5). Furthermore within countries with very large populations such as China, India, and the USA certain differences in the distribution of genotypes can be observed, linked to the country of origin of the carrier as well as their ethnicity (Sugauchi et al., 2003, Ding et al., 2002). Research has shown that Caucasians are more often infected with genotypes A and D, Asians with genotypes B and C (Westland et al., 2003, Chu et al., 2003), while genotype E has been predominately found in black HBV carriers (Lindh et al., 1997, Odemuyiwa et al., 2001, Westland et al., 2003). Comparable studies in Yemen found that natives were infected predominately with genotype D, and those of African origin with genotype A (Sallam and William Tong, 2004).

Serotypes

Within genotypes there is a great deal of diversity leading to the division of some genotypes into subtypes (Sugauchi et al., 2004a, Sugauchi et al., 2004b, Devesa et al., 2004, Huy et al., 2004, Norder et al., 2004). Different subtypes evolve
independently of others suggesting the origin of HBV subtypes is also geographical. Subtype A1 is most often found in Africa and Asia, A2 in North America and Europe (Sugauchi et al., 2004a), while B1 and B2 are found in Japan and the rest of Asia respectively (Sugauchi et al., 2004b). In South and East Asia, Sub-Saharan Africa and South and Central America populations show a particularly high frequency of HBV infection which can be maintained by vertical mother to child transmission or horizontal transmission during childhood (Dumpis et al., 2001). In sub Saharan Africa it is estimated that 5-10% of the adult population is chronically infected while rates of 2-5% are observed in the Middle East and Indian subcontinent (Andre, 2000, Gust, 1996). Less than 1% of the European and American population are chronically infected but approximately 5000 deaths per year are associated with HBV infection in the United States.

Figure 1.5: Geographical distribution of HBV genotypes and their relative prevalence in the host population. Based on the distribution of the 8 human genotypes, certain HBV genotypes can be found to predominate in specific geographical regions of the world. Recombinant C/D variants are denoted in striped purple. The numbers next to the pie charts are the number of isolates genotyped. Image reproduced with permission from the author (Kramvis et al., 2005).
HBV variants have further been detected in a divergent range of species including *Avihepadnaviruses* in herons, Ross’s geese, snow geese, storks, cranes and ducks (Sprengel *et al.*, 1988, Chang *et al.*, 1999, Pult *et al.*, 2001, Prassolov *et al.*, 2003, Mason *et al.*, 1980) and *Orthohepadnaviruses* in a woolly monkey (Lanford *et al.*, 1998), bats (Drexler *et al.*, 2013b), woodchucks (Cohen *et al.*, 1988), ground squirrels (Seeger *et al.*, 1984) and arctic squirrels (Testut *et al.*, 1996), non-human primates (MacDonald *et al.*, 2000, Hu *et al.*, 2001, Takahashi *et al.*, 2001, Vartanian *et al.*, 2002, Grethe *et al.*, 2000b, Sa-Nguanmoo *et al.*, 2009, Makuwa *et al.*, 2007) and more recently bats (Drexler *et al.*, 2013b). It was suggested that human HBV was acquired from NHPs or vice versa (Starkman *et al.*, 2003, Simmonds and Midgley, 2005). While several studies have been conducted to study cross-species transmission of human HBV to NHPs, infection of NHPs with human variants has not been documented in the wild and there is no supporting evidence for cross-species transmission of NHP HBV variants to humans.

### 1.1.6 HBV in NHP primates

Following the detection of HBV a number of studies in the 60’s and 70’s focused on screening NHPs for HBV infections. While the initial assays were crude the most consistent finding was data detailing the detection of HBsAg antibodies in 3-6% young chimpanzees (Eichberg and Kalter, 1980, Maynard *et al.*, 1971, Lichter, 1969, Lander *et al.*, 1972). The initial detection in NHPs was attributed to inoculation of these animals with human sera, although the subsequent detection and/or sequencing of HBV like viruses from a white handed gibbon (*Hylobates lar*) in Thailand (Gallagher *et al.*, 1991, Norder *et al.*, 1996) and in four of six woolly monkeys (*Lagothrix lagotricha*) in Louisville (Lanford *et al.*, 1998) suggested the presence of circulating HBV variants among NHPs distinct from human HBV.

As a consequence of these findings two retrospectives studies on archived chimpanzee sera and a further Cameroon study detected species-specific variants
At approximately the same time studies reported identifying HBV like genomes in gorillas, gibbons and orangutans (Warren et al., 1999, Lanford et al., 2000, Grethe et al., 2000b). Further studies have detected high frequencies of active and resolved infections in NHPs, and species-associated HBV variants distinct from human variants have now been detected in chimpanzees (Hu et al., 2001, Hu et al., 2000, MacDonald et al., 2000, Starkman et al., 2003, Takahashi et al., 2000, Makuwa et al., 2007, Vartanian et al., 2002), gorillas (Grethe et al., 2000b) and Southeast Asian gibbons and orangutans (Warren et al., 1998, Grethe et al., 2000b, Warren et al., 1999, Verschoor et al., 2001, Sall et al., 2005, Noppornpanth et al., 2003, Sa-Nguanmoo et al., 2009). The 10–12 NHP taxa-associated variants are distinct from the human variants of HBV and occur at infection frequencies comparable to human rates in endemic regions (Makuwa et al., 2005, Sa-Nguanmoo et al., 2009).

Hepadnaviruses of NHPs are indigenous to their hosts and phylogenetic analysis of complete genomes shows distinct clustering based on geographical origins (Grethe et al., 2000a, Norder et al., 1996, Robertson and Margolis, 2002, Sall et al., 2005, Warren et al., 1999, Starkman et al., 2003). HBV isolates from Pan troglodytes troglodytes (P.t.t) cluster closely with those from gorillas with whom they share overlapping geographical habitats. Similarly in island areas of South East Asia, sequences from orangutans cluster more closely with gibbons Hyllobates agilis, who are also found in this area while gibbon species of mainland South East Asia group separately. These studies together suggest a correlation between geographical clustering and the distribution of chimpanzee sub-species in Africa (Starkman et al., 2003, Makuwa et al., 2007), a finding further supported by the clustering in Asia of orangutan variants deep within the gibbon clade, where they occupy overlapping and adjacent habitats in contrast to variants infecting Hyllobates lar (H. lar) and mainland Asia gibbons which form separate groups (Starkman et al., 2003, Norder et al., 1996, Sall et al., 2005). There is preliminary evidence to suggest that viral sequences from
Introduction

geographically distinct sub-species of chimpanzee form distinct phylogenetic clusters supporting the hypothesis that HBV co-speciated within chimpanzees (Hu et al., 2001). The suggestion that the introduction of HBV into the orangutan species is a relatively recent event is supported by the lack of sequence diversity observed between isolates and also the narrow geographical distribution of the virus within the species (Starkman et al., 2003, Sall et al., 2005, Warren et al., 1999, Warren et al., 1998, Sa-Nguanmoo et al., 2009).

Examination of the nucleotide divergence observed between HBV isolates from African and Asian apes reveal 10-11% rate of divergence, while rates of between 5-7% and 7-9% are observed between gorilla and chimpanzee and gibbon and orangutan isolates, respectively. With the exception of the woolly monkey isolate which is 28% divergent from human HBV variants; all other NHP isolates display 10-15% nucleotide divergence (Lanford et al., 1998).

1.1.7 Recombination within and between HBV variants infecting humans and non-human primates

Recombination plays a major role in the evolution of viruses and the generation of novel variants and improving the fitness of the viral population, increasing infectivity, pathogenicity, or host range. These recombinants can generate novel HBV variants, contributing to the genetic diversity of the virus. Recombination was previously identified in HBV, a 196bp region of the preCore/Core region was shown to enhance recombination in vitro in the presence of extracts from actively dividing cells (Hino et al., 1991). Studies looking at integrated and episomal HBV in a single patient also detected indications for homologous recombination (Georgi-Geisberger et al., 1992). Subsequent studies applied phylogenetic analysis to available HBV sequences identifying recombination in 2/25 and 14/65 samples between genotypes A/D and B/C in two separate studies (Bollyky et al., 1996, Bowyer and Sim, 2000), findings supported by future studies also identifying A/D (Morozov et al., 2000,
Owiredu *et al.*, 2001b) and B/C recombinants (Bollyky *et al.*, 1996, Morozov *et al.*, 2000, Sugauchi *et al.*, 2002b, Mukaide *et al.*, 1992). Incongruences observed in the phylograms generated using different ORFs of the HBV genome suggested recombination was a factor in the evolution of the virus, and identified a recombinant C/D variant in Tibet and C/A variant in East Asia (Mizokami *et al.*, 1994, Norder *et al.*, 1994, Sugauchi *et al.*, 2002a, Hannoun *et al.*, 2000b, Cui *et al.*, 2002). Recent research has further detected a novel recombinant between genotypes H/B in a Japanese patient (Uchida *et al.*, 2013). A large study by Simmonds and Midgley (Simmonds and Midgley, 2005) applied a novel tree order scan technique to determine the level of recombination between HBV genotypes globally. Results found that 90% of recombinants identified were B/C or A/D hybrids. The remaining recombinants were A/B/C, A/C, A/E, A/G, C/D, C/F, C/G, C/U and B/C/U were ‘U’ is an unknown genotype. Further this study identified recombination breakpoints in the HBV genome concentrated around the DR1 and pre-S1/S2 regions, and at the 3’ end of the core and surface genes.

Certain HBV hybrids have become the predominantly circulating variants, for example subgenotype B2 recombined with genotype C in the pre-core/core regions predominates in Asia, while in Japan the non-recombinant subgenotype B1 predominates (Sugauchi *et al.*, 2002a). In Tibet the recombinant C/D variant is responsible for 96% of HBV infections (Cui *et al.*, 2002, Wang *et al.*, 2005, Zeng *et al.*, 2005).

The occurrence of recombination between human and NHP HBV variants has been documented. Recombination events have been detected between the human genotype C and the chimpanzee (*Pan troglodytes schweinfurthii*) variant AF498266 (Magiorkinis *et al.*, 2005) and between human and gibbon/orangutan variants (Sa-Nguanmoo *et al.*, 2009, Tatematsu *et al.*, 2009). Their occurrence demonstrates that despite their genetic divergence, human and non-human associated
variants of HBV can share hosts in nature. Certain questions still remain surrounding HBV and the role HBV recombination may play, for example, in the geographical differences observed in pathogenicity and viral evolution and the variations in immune responses observed in infected patients (Foster and Thomas, 1993). A recently published study characterising HBV variants infecting ape populations in Cameroon (Njouom et al., 2010) demonstrated the existence of a gorilla-specific HBV strain and evidence of recombination between HBV strains circulating in chimpanzees. This and previous studies of HBV nucleotide sequence similarity indicate non-human primates (NHP) have distinct species-specific variants of HBV distinguishable both from each other and from human HBV despite occupying overlapping geographical areas (MacDonald et al., 2000, Robertson and Margolis, 2002, Norder et al., 1996).

1.1.8 Proposed models and hypotheses of HBV evolution

It has been difficult to propose potential evolutionary hypotheses which may account for the interspersed sequence relationships between the human HBV genotypes and those infecting gorillas, and chimpanzees in Africa and gibbons and orangutans in South Asia. A primate origin for HBV has been suggested. Initial estimates of a mean substitution rate of $4.2 \times 10^{-5}$ nucleotide substitutions per site per year, implied that HBV in humans and NHPs occurred approximately 6000 years ago (Fares and Holmes, 2002). Applying a higher substitution rate between $1.5-2.1 \times 10^{-5}$ substitutions per site per year suggests the human HBV genotypes originated from a common ancestor 2300-3100 years ago (Okamoto et al., 1987, Hannoun et al., 2000a, Simmonds, 2001a). This finding is concordant with another study estimating that human and NHP HBV variants diverged 3000 years ago and implies subsequent evolution of the virus in a host dependent manner (Orito et al., 1989).

Due to the compact genomic organisation of HBV with several overlapping reading frames, limitations are placed on our ability to model constraints on sequence change that might effect the long term diversification of the HBV. Three models have been
proposed to explain the genetic evolution of HBV in humans, non-human primates and other species.

‘Out of South America’

Prior to the detection of NHP HBV, an initial hypothesis based on phylogenetic analysis of human HBV isolates proposed HBV originated in the Americas and as a consequence of colonisation spread to the Old World about 400 years ago (Bollyky et al., 1997). Similar conclusions were observed when examining WMHBV sequence that was suggested to be the origin of human HBV (Lanford et al., 1998). Such a hypothesis would subsequently have meant that humans were the source of NHP infections over a relatively short period of time (Simmonds, 2001b). Following the detection of infections in wild primates in Africa and Asia it has been necessary to revise this hypothesis.

‘Out of Africa’

Studies have examined the potential that HBV evolution corresponds to the dissemination of humans out of Africa between 100,000 and 200,000 years ago (Norder et al., 1994, Magnus and Norder, 1995). This model is consistent with the detection in Australian aborigines of genotype D and C, assuming that these genotypes were present in the populations that migrated out of Southeast Asia over 40,000 years ago (Sugauchi et al., 2001). This proposal would further require that the genetic relationships between human populations be reflected in the phylogeny of HBV genotypes, as is the case for HPgV (Pavesi, 2001) and this is not observed for all HBV genotypes. Genotype B and C infecting populations of South and East Asia for example, is inconsistent with the presence of genotype F in Native Americans, their nearest relatives (Alestig et al., 2001, Arauz-Ruiz et al., 2002). Furthermore the theory would require that HBV in NHPs display far greater genetic diversity between species and between human and NHP HBV given the duration of co-speciation of primate species (≈15 million years). This would require a divergence greater than the
11% currently estimated between human and chimpanzee-HBV variants (MacDonald et al., 2000).

Co-speciation

The third hypothesis argues that HBV variants in gorillas, chimpanzees, gibbons, orangutans and the Woolly monkey co-speciated within primates between 10-35 millions years ago following cross-species transmission (MacDonald et al., 2000, Simmonds, 2001b). Such a hypothesis is supported by the finding of species-specific variants of HBV in diverse species such as squirrels and woodchucks that may be the result of comparable coevolution events over extended periods of time. It is further supported by the detection of chimpanzee sub-species specific HBV variants in different geographical locations and the formation of distinct gorilla and chimpanzee sub-species clades of HBV when analysed by phylogenetic analysis (Hu et al., 2001, Njouom et al., 2010). Geographical factors also appear to play an important role, for example the phylogenetic clustering of orangutan HBV variants from Island Borneo within the gibbon clade while other mainland gibbon variants from other regions of South East Asia form distinct clades (Noppompanth et al., 2003). The co-speciation model supports the outlier position of the WMHBV variant and the equal divergence observed between HBV variants among NHPs.

Again there are certain limitations to the co-evolution hypothesis, as it would require a far lower rate of long-term sequence evolution of HBV between 3-5 x10⁹ nucleotide changes per site per year. The model does not explain the presence of the equally distinct human genotypes, unless comparable to the spread of HIV-1 and HIV-2 to humans from chimpanzees and sooty mangebys respectively (Gao et al., 1999, Feng et al., 1992), that the observed HBV genotypes are also the result of several cross-species transmission events. The observation that areas of high HBV prevalence in the human population are also those areas in which humans and NHP populations share habitat or are in regular contact, supports a primate origin for HBV. However, to date no wild primates have been found infected with human
variants, or vice versa, therefore no cross-species transmission events have been confirmed. The recent detection of three novel species of hepadnavirus infecting bats in Central America and Africa has given rise to the suggestion that bats may be the natural reservoirs for hepadnaviruses and were subsequently transmitted to primates/humans through the hunting and consumption of bush meat (Drexler et al., 2013b).

Cameroon is within a region of endemic human HBV infection with a hepatitis B surface antigen (HBsAg) prevalence in humans of 8% or greater (Mbanya et al., 2003). Additionally, four different great ape taxa also occur in Cameroon, providing the conditions for potential inter-species transmission. Therefore the potential for cross-species transmission to occur is much higher in these regions and the probability of detecting such an event is greater. Samples from NHPs obtained for this study all originated in different areas of Cameroon and aimed to establish whether these primates were host to human HBV variants. Although no human-derived genotypes of HBV were detected in non-human primates screened in the study presented here, evidence for transmission of HBV between chimpanzee subspecies and between chimpanzees and gorillas was obtained. Host mitochondrial sequencing and phylogenetic analysis of HBV variants from chimpanzees was applied to find supporting evidence for co-speciation or allopatric speciation of HBV among non-human primates in Cameroon.

1.1.9 Mechanisms of HBV recombination

Recombination is a process that can lead to large evolutionary jumps for viruses and substantially improve the fitness of the viral population, potentially increasing infectivity, pathogenesis and host range. In the context of the viral life-cycle of HBV, there are multiple steps when recombination events could occur, including RNA–RNA during packaging of pgRNA as well as DNA–DNA during reverse transcription and generation of double-stranded linear genomes (Yang and Summers,
1995), and finally, during transcription of viral RNA from the different sub-populations of viral minichromosomes (Newbold et al., 1995). The exact point at which recombination occurs during HBV replication has not been determined. Opportunities for recombination are present during minus and plus-strand synthesis. The special structure of rcDNA could lend itself to recombination events. Recombination is unlikely to occur during the reverse transcription of pgRNA as this occurs following encapsidation of only a single pgRNA. Individuals and animals co-infected with two different HBV genotypes can result in the exchange of genetic material between the two viral strains by homologous recombination, the exchange of nucleotide sequences between two similar or identical molecules of DNA. Therefore it is more likely that recombination events occur in the nucleus where segments of the genome can be exchanged between cccDNA molecules derived from hosts co-infected with different genotypes. Such events are probably rare; although this process of recombination has the advantage in that the generated recombinant can be immediately produced by pgRNA transcription from the chimeric cccDNA, explaining the more widespread detection of HBV recombinants.

Homologous recombination was first documented in hepadnaviruses by in vivo DNA transfection using duck HBV (DHV) (Sprengel et al., 1987). A 196bp region of the HBV pre-core/core region has been demonstrated to enhance recombination in the presence of actively dividing cells (Hino et al., 1991). A hepatocellular carcinoma has been shown to carry HBV recombinants generated by homologous recombination (Magiorkinis et al., 2005). Subsequently two studies applied phylogenetic analyses of HBV sequences to detect the presence of HBV recombination in 2/25 and 14/65 samples respectively (Bowyer and Sim, 2000, Bollyky et al., 1996).

Recombination between genotypes occurs opportunistically in geographical regions where several genotypes co-circulate. As previously discussed in Section 1.2.7
genotype A and D recombinants have been found in Africa (Owiredu et al., 2001a), whereas genotype B/C recombinants occur in Asia (Sugauchi et al., 2002a, Bowyer and Sim, 2000) and C/D in Tibet (Cui et al., 2002). Genome regions containing topoisomerase I cleavage sites display properties favouring recombination (Kew et al., 1993, Konopka, 1998, Bullock et al., 1985). In vitro studies of DHV have shown that topoisomerase I mediates cleavage and recombination (Pourquier et al., 1999). Overlapping regions of the DR1 and DR2 have high concentrations of topoisomerase I sites, creating hot spots for recombination (Shih et al., 1987). Recent studies have reviewed the level of HBV recombination Tree Order Scanning and breakpoint analysis (Simmonds and Midgley, 2005, Yang et al., 2006). As mentioned above recombinants were detected between B/C or A/D in 90% of the sequences analysed, in addition to recombinants detected between A/B/C, A/C, A/E, A/G, C/D, C/F, C/G, C/U (U=unknown genotype) and B/C/U hybrids. Theses studies detected recombination breakpoints concentrated around the DR-1 region (1640–1900nt), the 3’-end of the core gene (2330–2450nt) and surface gene (650–830nt), and the pre-S1/S2 region (3150–100nt).

1.2 Hepatic viruses of the Flaviviridae family

1.2.1 Flaviviridae family and classification

The family of Flaviviridae is currently divided into the 4 genera of Flavivirus, Hepacivirus, Pestivirus and Pegivirus (Figure 1.6). Members of this family are small, enveloped viruses and contain a positive-sense, ssRNA genome, which encodes a single polyprotein (Thiel et al., 2005). The Pegivirus genus is a recently designated genus and includes the previously unassigned viruses human pegivirus [HPgV] (GB Virus C (GBV-C)) and simian pegivirus [SPgV_cpz, SPgV] (formerly GBV-A) and the recently identified equine, rodent and bat pegiviruses, EPgV and TDAV, RPgV and BPgV pegivirus respectively (Kapoor et al., 2013a, Kapoor et al., 2013b, Quan et al., 2013, Stapleton et al., 2011). Until recently HCV and GBV-B were the only
members of the genus *Hepacivirus*, which now includes homologs identified in horses (nonprimate hepacivirus [NPHV]) rodents (rodent hepacivirus [RHV]) and bats (bat hepacivirus [BHV]). SPgV and GBV-B were discovered in tamarins infected with the sera of a patient (initials GB) with mild hepatitis and are closely related to HCV (Deinhardt *et al.*, 1967, Simons *et al.*, 1995a, Simons *et al.*, 1995b). NPHV is the phylogenetically closest relative to HCV (Kapoor *et al.*, 2013b, Quan *et al.*, 2013). GBV-B infected tamarins develop an acute infection that evolves to chronicity similar to HCV infection (Bukh *et al.*, 1999, Martin *et al.*, 2003), whether this occurs for NPHV, RHV and BHV has yet to be determined. HPgV is closely related to SPgV and GBV-B and shares about 25-30% sequence similarity with HCV (Leary *et al.*, 1996b). HPgV infects humans, but has not been associated with any disease. However, it is believed that HPgV can prolong HIV disease progression (Shankar *et al.*, 2008).
Figure 1.6: Current genera of the Flaviviridae family and their respective members. This excludes the more recent as yet unassigned isolates from rodents (RHV/RPgV), bats (BHV/BPgV) and GHV from a new world monkey. Phylogenetic analysis was carried out on the conserved sequences in the RdRp of classified members of the Flaviviridae family. An unrooted phylogenetic tree constructed from the sequence alignment by neighbour-joining of (uncorrected) amino acid p-distances using the MEGA version 4.1 package. (Image reproduced with permission from authors (Simmonds, 2013)).
1.2.2 Genome organisation of Hepacivirus and Pegivirus genus

Hepacivirus-HCV Genome

The HCV genome is 9.6 kb long and translated into a single long polyprotein that subsequently undergoes co-translational and post-translational proteolytic processing in the cytoplasm or ER of the infected cell. Host and viral peptidases process the translated polyprotein from the N-terminal region into 10 structural and non-structural proteins: core, E1, E2, p7, NS2, NS3, NS4A, NS4B, NS5A and NS5B (Xu et al., 2001, McMullan et al., 2007, Walewski et al., 2001). The genome is structured into the N-terminal domain (core-E2) containing the structural proteins and the remaining C-terminal two-thirds of the genome containing the non-structural proteins (NS2-NS5B)(Figure 1.7).

![Genome organisation of members of the Hepacivirus genus.](image)

Figure 1.7: Genome organisation of members of the Hepacivirus genus. Genomes vary in length from 9.2kb-11.2kb. HCV homologs and genetically related viruses have been detected in primates (GBV-B), rodents (RHV), bats (BHV), horses (NPHV), New World monkeys (GHV) and possibly dogs. Structural proteins are marked in blue, and non-structural in yellow. Enzymes responsible for the Proteolytic procession of the polyprotein are depicted with black, red and blue arrows.

Different regions of the HCV genome display high degrees of sequence variability
Introduction

(Martell et al., 1992, Pawlotsky, 2006). The structural regions of E1 and E2 display the highest level of variability, while the structural 5’UTR and 3’UTR and genetically conserved NS3 helicase and NS5B RNA polymerase show high levels of sequence conservation among HCV variants. The 5’UTR of HCV is approximately 341nt and GBV-B 445nt in length and contain an internal ribosomal entry site (IRES) with four highly structured domains I-IV. The 5’UTR and the contained IRES directs the translation of a core protein. The 5’UTR IRES is essential for cap-independent translation of the HCV viral RNA and is mostly conserved between HCV and genetically related viruses (Bukh et al., 1992, Brown et al., 1992, Honda et al., 1999, Tsukiyama Kohara et al., 1992, Wang et al., 1993). As seen in other viruses the HCV 5’UTR contains cis-acting replication elements (CREs), defined stem-loop structures that can be found in the 5’UTR, core, NS5B and 3’UTR that are cis-acting and an essential component of RNA replication. These CRE elements can additionally form interactions in long-range kissing-loops further regulating viral replication (Friebe et al., 2001, Kim et al., 2003, You et al., 2004, Diviney et al., 2008, Romero-Lopez and Berzal-Herranz, 2009). Mutations in the CRE element of the NS5B (stem-loop 5BSL3.2) have been found to block RNA replication and it is thought that both the primary sequence and the structure of this element are crucial for viral replication (You et al., 2004). The 5’UTR additionally contains a mir122-binding site in the IRES that is capable of binding to the liver specific microRNA-122 enhancing replication (Jopling et al., 2005). The first 125nt of the HCV 5’UTR are sufficient for RNA replication, although the sequences upstream of the IRES, in addition to stem-loop II (SLII) of the IRES are essential for efficient viral replication (Friebe et al., 2001).

The HCV 3’UTR varies in length from 200-235nt and includes a short variable region, a poly (U/UC) tract with an average length of 80nt and a nearly invariant 98nt long X-tail region (Kolykhalov et al., 1996, Tanaka et al., 1995, Tanaka et al., 1996). The 3’X-tail forms 3 stable stem-loop structures, also called “clover-leaf”,

28
that are highly conserved among all genotypes and are essential for RNA replication. The 3’X region and the 52nt upstream of the poly (U/UC) tract are crucial for RNA replication, while the rest of the 3’UTR plays a role in the enhancement of replication (Yi and Lemon, 2003, Friebe and Bartenschlager, 2002).

**Table 1.1: Estimated length of the 5’UTR and 3’UTR of members of the Hepacivirus genus.**

Numbers are estimates based on available sequences. ND represents data that is not determined.

<table>
<thead>
<tr>
<th>Virus</th>
<th>5’UTR</th>
<th>3’UTR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hepacivirus Genus</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HCV</td>
<td>~341bp</td>
<td>~230bp</td>
</tr>
<tr>
<td>GBV-B</td>
<td>~445bp</td>
<td>~350bp</td>
</tr>
<tr>
<td>RHV</td>
<td>~550bp</td>
<td>~230bp</td>
</tr>
<tr>
<td>BHV</td>
<td>~280bp</td>
<td>~250bp</td>
</tr>
<tr>
<td>NPHV</td>
<td>~390bp</td>
<td>ND</td>
</tr>
</tbody>
</table>

**Pegivirus**

Like HCV, members of the *Pegivirus* genus (SPgV, HPgV, BPgV, EPgV, TDAV and RPgV) have a positive-sense, ssRNA genome that contains a single long ORF encoding a multifunctional polyprotein (Figure 1.8) (Birkenmeyer et al., 1998, Choo et al., 1991, Kim and Fry, 1997, Leary et al., 1996b, Muerhoff et al., 1995). The pegivirus genome varies in length from between 8.9kb for BPgV to 11.2kb for RPgV and EPgV, with HPgV and SPgV being intermediate at 9.4 kb. Pegivirus genomes are predicted to contain 5’ untranslated region (5’UTR) and 3’UTR and contain a long open reading frame (ORF) encoding approximately 3000 amino acids that are post-translationally cleaved into structural and non-structural proteins. The viral structural proteins E1 and E2 are processed from the amino-terminal portion of the polyprotein by cellular signal peptidases or signal peptide peptidase, while the NS proteins NS2, NS3, NS4, NS5A and NS5B in the remaining two thirds of the genome are processed by two viral encoded proteases (Penin et al., 2001, Kapoor et al., 2013a, Kapoor et al., 2013b, Quan et al., 2013, Epstein et al., 2010). It is predicted that there is an additional glycoprotein between the E2 and NS2 of BPgV, EPgV an
Introduction

RPgV. This region of the polyprotein is called the ‘X’ protein (Kapoor et al., 2013a, Kapoor et al., 2013b, Epstein et al., 2010). It is not known whether SPgV and HPgV have a corresponding protein. Furthermore pegiviruses do not appear to encode a core protein (Kapoor et al., 2013a, Kapoor et al., 2013b, Kim and Fry, 1997, Leary et al., 1996b, Xiang et al., 1998, Epstein et al., 2010). The 5’UTR contains an internal ribosomal entry site (IRES) element that directs translation of the polyprotein directly from viral genomic RNA (Simons et al., 1996, Yoo et al., 1992), and translation and release of the viral RNA-dependent RNA polymerase (RdRp) initiates RNA replication (Major and Feinstone, 1997, Mohr and Stapleton, 2009, Moradpour et al., 2007, Robertson, 2001).

**Figure 1.8: Genome organisation of members of the Pegivirus genus.** Genomes vary in length from 9.2kb-11.2kb. HPgV homologs have been detected in a range of species including bats (BPgV), rodents (RPgV), horses (EPgV and TDAV), and previously in primates (SPgVcpz/SPgVtrog). Structural proteins are marked in blue, and non-structural in yellow. Enzymes responsible for the Proteolytic procession of the polyprotein are depicted with black, red and blue arrows.
The predicted 5′UTR of HPgV (555bp), RPgV (349bp) and EPgV (615bp) contains an internal ribosome entry site (IRES) that directs translation of the mRNA (Simons et al., 1996, Kapoor et al., 2013a, Kapoor et al., 2013b), while there is currently insufficient evidence to determine the presence of an IRES element in the TDAV 5′UTR (617bp) (Chandriani et al., 2013). All of the pegiviruses studied have an IRES element in the 5′UTR, the IRES elements in SPgV and HPgV conform best to type 4 IRES elements (Kieft, 2008), while IRES structures for EPgV predict structural elements comparable to the type 1 IRES of enteroviruses (Kapoor et al., 2013a). However, others state that the SPgV and HPgV IRES do not conform to any recognized IRES class (Bakhshesh et al., 2008) and IRES activity has not yet been examined for BPgV.

The HPgV (~300bp), EPgV (664bp) and RPgV (475bp) 3′UTRs are different from that of HCV and GBV-C in that they do not contain poly-(A) or poly-(U) tracts, though it shares predicted structural elements (Xiang et al., 2000, Kapoor et al., 2013a, Kapoor et al., 2013b). Analysis of the 3′UTR of EPgV determined the presence of 3-6 repeat sequence elements (RSEs) followed by a highly conserved 380bp sequence (Kapoor et al., 2013a). The RSEs of EPgV were further predicted to fold into conserved stem-loop structures not previously described in other mammalian viruses. The 3′UTR of RPgV contains two poly-C tracts of 9nt and further stem-loop structures downstream of the stop codon (Kapoor et al., 2013b).
Table 1.2: Published estimates of the 5’UTR and 3’UTR of identified members of the Pegivirus genus. All figures are estimates based on currently available sequences.

<table>
<thead>
<tr>
<th>Virus</th>
<th>5’UTR</th>
<th>3’UTR</th>
</tr>
</thead>
<tbody>
<tr>
<td>HPgV</td>
<td>~555bp</td>
<td>~300bp</td>
</tr>
<tr>
<td>SPgVcpz</td>
<td>~550bp</td>
<td>~300bp</td>
</tr>
<tr>
<td>SPgV</td>
<td>~540bp</td>
<td>~100bp</td>
</tr>
<tr>
<td>EPGV</td>
<td>615bp</td>
<td>~664bp</td>
</tr>
<tr>
<td>RPGV</td>
<td>~350bp</td>
<td>~475bp</td>
</tr>
<tr>
<td>BPGV</td>
<td>~290bp</td>
<td>~400bp</td>
</tr>
</tbody>
</table>

1.2.3 Proteolytic cleavage of HCV and HPgV and related viruses

Processing of the HCV polyprotein has been experimentally documented, although the processing of members of the Pegivirus genus has largely been based on sequence comparison with HCV. Cellular proteases process the junctions between core/E1, E1/E2, E2/p7 and p7/NS2 followed by the intramolecular cleavage of junction NS2/NS3 by the NS2/NS3 autoprotease. Downstream, intramolecular cleavage processes the junctions between NS4A/NS4B/NS5A/NS5B and intermolecular cleavage the NS3/NS4A both by the action of the NS3/NS4A protease (Figure 1.7, 1.8).

Comparable to HCV a cellular signal peptidase is predicted to cleave pegivirus junctions E1-E2, E2-P7 and P7-NS2, based on consensus eukaryotic signal sequence cleavage sites (Leary et al., 1996b). Based on the conservation of His-952 and Cyc-993 at the NS2 junction as seen in HCV, the NS2 viral serine protease is predicted to mediate cleavage at NS2/NS3 in pegiviruses (Belyaev et al., 1998, Kapoor et al., 2013a, Kapoor et al., 2013b, Stapleton et al., 2011) (Figure 1.8). Cleavage of the remaining NS proteins NS3/NS4 and NS5A/NS5B is mediated by the NS3 protease and, together with NS4A, mediates subsequent cleavage of the NS4B/NS5A junction (Belyaev et al., 1998, Stapleton et al., 2011, Kapoor et al., 2013a, Kapoor et al., 2013a, Kapoor et al., 2013b, Stapleton et al., 2011).
The NS3 protease is thought to mediate the cleavage of NS4A/NS4B, as it predicted to mediate this cleavage in HCV (Belyaev et al., 1998).

1.2.4 Functions of encoded viral proteins

The hepacivirus and pegivirus open reading frame (ORF) encode at least 11 proteins, including structural proteins and non-structural, whose functions are continuously being studied and elucidated.

**Table 1.3: Generalised functions of the hepacivirus and pegivirus encoded proteins in the viral life cycle.**

<table>
<thead>
<tr>
<th>Protein</th>
<th>Role</th>
<th>Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>Core</td>
<td>Structural</td>
<td>ER localisation/ Nucleocapsid formation</td>
</tr>
<tr>
<td>E1</td>
<td>Structural</td>
<td>Envelope/ Fusion domain</td>
</tr>
<tr>
<td>E2</td>
<td>Structural</td>
<td>Envelope/ Binding to cell receptor/ Fusion domain</td>
</tr>
<tr>
<td>P7/X</td>
<td>Structural</td>
<td>Ion Channel</td>
</tr>
<tr>
<td>NS2</td>
<td>Non- Structural</td>
<td>Autoprotease NS2/NS3</td>
</tr>
<tr>
<td>NS3</td>
<td>Non- Structural</td>
<td>NTPase/helishe/ NS2-NS3 autoprotease</td>
</tr>
<tr>
<td>NS4A</td>
<td>Non- Structural</td>
<td>Cofactor of NS3/NS4A proteinase</td>
</tr>
<tr>
<td>NS4B</td>
<td>Non- Structural</td>
<td>Membranous web formation</td>
</tr>
<tr>
<td>NS5A</td>
<td>Non- Structural</td>
<td>Replication complex formation</td>
</tr>
<tr>
<td>NS5B</td>
<td>Non- Structural</td>
<td>RNA-dependent RNA polymerase</td>
</tr>
</tbody>
</table>

Core

The core structural protein is an RNA binding protein that forms the nucleocapsid. The presence of an internal signal sequence between the core and E1 proteins targets the newly synthesized polyprotein to the endoplasmic reticulum (ER), where E1 is translocated into the ER, and cleaved at position 191 by the ER signal peptidase, producing an immature form of the core (Santolini et al., 1994, Yasui et al., 1998). The mature form of the core is released into the cytosol by further cleavage of the signal peptide by the signal peptide peptidase (McLauchlan et al., 2002, Hussy et al., 1996, Santolini et al., 1994). The mature core is transferred from the ER membrane
to the surface of lipid droplets (LDs), where it interacts with NS5A also attached to LDs (McLauchlan et al., 2002, Barba et al., 1997). Core is a multifunctional protein essential for viral replication, maturation and pathogenesis. In addition to its role in the formation of the HCV virion the protein also shows some modulatory functions, in regulating signalling pathways, viral and cellular gene expression, lipid metabolism, apoptosis, and cell transformation (McLauchlan, 2000, Suzuki and Suzuki, 2006) and induces hepatic steatosis in transgenic mice (Moriya et al., 1997).

**E1 and E2 envelope proteins**

E1 and E2 are glycosylated envelope proteins required for virus entry (Helle et al., 2006, Goffard et al., 2005). They are type-I trans-membrane proteins (TM), are N-glycosylated in the N-terminal ectodomains and anchored into membranes by the hydrophobic C-terminal transmembrane domains (TMDs) (Op et al., 2001). The C-terminal TMDs function as membrane anchors, ER retention signals and are involved in E1/E2 heterodimerisation (Cocquerel et al., 2002). In addition to the E1/E2 heterodimers stabilized by non-covalent interactions and inserted into the viral envelope, E1 and E2 can form additional misfolded di-sulphide aggregates (Dubuisson et al., 1994). Based on comparison with HCV, it is thought that the pegivirus ectodomains are also directed to the ER lumen and the TMDs are inserted into the ER membrane becoming type-I transmembrane proteins (Kaiser and Tillmann, 2005). The E1/E2 heterodimer of HCV and pegiviruses are predicted to be retained in the ER due to the presence in the E2 of the N-linked oligosaccharides (Pilot-Matias et al., 1996a). Some HCV heterodimers can be further detected on cellular plasma membranes when the envelope glycoproteins are overexpressed (Dubuisson, 2007).

E1 and E2 are also the gene regions with the highest genetic diversity in the HCV genome. E2 contains hypervariable domains (e.g., HVR1) that are the most variable regions of the HCV genome (Hijikata et al., 1991, Weiner et al., 1991, Kato et al.,
1992). These hypervariable regions have not been detected for HPgV or other pegiviruses (Stapleton et al., 2004), which may prevent the virus from producing neutralization escape variants and subsequently may contribute to the increased rate of viral clearance observed in infected individuals. The high level of variability of in the HCV HVR1 is potentially involved in HCV persistence, allowing the virus to escape the host immune responses (Kumar et al., 1994).

p7 protein
The p7 protein found in HCV and GBV-B is a small hydrophobic polypeptide that spans the ER double membrane and is an essential component of virus assembly and release and the production of infectious virions in vivo (Sakai et al., 2003, Brohm et al., 2009, Lindenbach and Rice, 2005). The p7 functions as an ion channel and has been classified as a viroporin, as it is a small channel-forming viral membrane protein that affects the infectivity of the virus and can be blocked by amantadine (Gonzalez and Carrasco, 2003, Griffin et al., 2003). The ion channels formed by p7 may play an important role in the life-cycle of HCV and drugs that block them may affect virus replication (StGelais et al., 2009). Comparative studies of pegiviruses with HCV identified a putative amphipathic protein at the C-terminus of E2 analogous to the HCV P7 protein. Two putative signal peptidase sites were identified surrounding an amino acid that may encode a small 5.6kD amphipathic helical protein that may serve as an HCV P7 homolog (Epstein et al., 2010).

NS2 protein
NS2 is a transmembrane protein, functioning as a highly hydrophobic protease responsible for the cleavage of the NS2/NS3 in association with the N terminus of NS3. The NS2 N terminus is highly hydrophobic and forms 3-4 transmembrane helices in the ER membrane and the NS2/NS3 junction is subsequently cleaved by zinc-stimulated cysteine protease, for which the NS3 N terminus functions as a cofactor (Schregel et al., 2009). It has also been suggested that NS2 plays an
important role in viral assembly and release (Pietschmann et al., 2006). The N-terminus (810-940) has been further shown to modulate host cell gene expression and apoptosis (Dumoulin et al., 2003, Machida et al., 2001, Erdtmann et al., 2003). Recently it has been shown that NS2 plays a role involved in the inhibition of cyclophilin A (CypA), a NS5B modulator (Ciesek et al., 2009).

**NS3/4A complex**

NS3-4A serine protease consists of catalytic subunit the N-terminal serine protease and an activating cofactor, NS4A, a 54-residue amphipathic peptide with a hydrophilic N-terminus and a hydrophobic C-terminus (Kim et al., 1996, Failla et al., 1994). NS3 is a multi-functional protein with an additional C-terminal region predicted to have RNA helicase/NTPase activity. The NS3-4A complex is responsible for cleavage of the four junctions NS3/4A (self cleavage), NS4B/NS5A and NS5A/NS5B (Bartenschlager et al., 1993, Bartenschlager et al., 1995, Failla et al., 1995, Grakoui et al., 1993a, Grakoui et al., 1993b, Hijikata et al., 1993, Kim et al., 1996, Tanji et al., 1995, Tomei et al., 1993). The NS3 protease also cleaves in trans within the NS3 helicase domain, with the NS3 helicase and NS4A as cofactors. The amino acid residues of NS4A involved in the internal cleavage of NS3 are different from those necessary for the cofactor activity of the NS3 serine protease (Shoji et al., 1999, Yang et al., 2000, Pan et al., 2009, Kou et al., 2007).

NS3/4A is furthermore involved in the blocking of the host cell innate immune response (Foy et al., 2003). It has been shown to interfere with the double-stranded RNA (dsRNA) signalling pathway, suppressing the interferon (IFN) induction in HCV replicating cells. The NS3 C-terminus encodes a DexH/D-box RNA helicase responsible for unwinding of the RNA and it can be structurally divided into an RNA nucleoside triphosphatases (NTPase) domain, RNA binding domain and a helical domain. Members of the DexH/D-box helicase superfamily 2 unwind RNA-RNA substrates in a 3'- to -5' direction and fuelled by ATP hydrolysis, this movement
allows the protein to displace complementary strands of DNA or RNA and proteins bound to the nucleic acid (Tai et al., 1996, Cho et al., 1998). Structural analysis has suggested that the NS3 helicase NTPase domain forms a dimer with the RNA binding domain of another helicase (Cho et al., 1998). The helicase is involved in the initiation of RNA replication combined with NS5A and NS5B (Zhong et al., 2005, Murayama et al., 2007). The helicase activity is positively modulated by the NS3 protease domain and NS4A (Frick et al., 2004).

**NS4B protein**

NS4B is a 27kDa hydrophobic protein which for many years was a relatively poorly characterised integral membrane protein, predicted to contain at least 4 transmembrane segments, a cytosolic N-terminal part and a cytosolic C-terminal involved in mediating membrane association (Lundin et al., 2003, Gouttenoire et al., 2009, Gosert et al., 2003). The amphipathic α-helix is also involved in the induction of the membranous web, which serves as a scaffold for the RNA replication complex (Egger et al., 2002). Similar to other HCV NS proteins, NS4B has been reported to play a role in virus assembly and release (Jones et al., 2009) and has been implicated in modulation of NS5B's RNA dependent RNA polymerase (RdRp) activity (Piccininni et al., 2002) and various host signal transduction pathways, a possible role in HCV carcinogenesis, impairment of ER function, and regulation of both viral and host translation (Hugle et al., 2001, Kato et al., 2002b, Konan et al., 2003, Park et al., 2000). Furthermore, NS4B has been demonstrated to have an NTPase activity and to play a role in HCV pathogenesis (Thompson et al., 2009, Gouttenoire et al., 2010).

**NS5A protein**

The NS5A protein has generated wide interest because of its ability to modulate the host-cell IFN response. It is a membrane anchored zinc-metalloprotein that exists in a basally (56 kDa) and hyperphosphorylated (58 kDa) form (Tellinghuisen et al., 2004,
Tellinghuisen et al., 2005). No enzymatic function has been assigned to NS5A but it has been shown to be essential for HCV replication and has significant impact on cellular pathways and processes, including the growth of host cells and proliferation. NS5A hyperphosphorylation inhibits HCV replication (Tellinghuisen et al., 2005, Evans et al., 2004, Blight et al., 2000, Krieger et al., 2001) and has been shown to interact with NS5B RNA dependent RNA polymerase (RdRp), an essential interaction for HCV RNA replication (Shimakami et al., 2004, Shirota et al., 2002). NS5A binds to 3'-ends of HCV plus and minus strand RNAs with high affinity and is important for genome replication (Huang et al., 2005). Recently it was found that domain III of NS5A is implicated in both RNA replication and assembly of HCV particles in JFH1-infected cells (Hughes et al., 2009). It also plays a crucial role in evading the host immune system through inhibition of the IFN-induced dsRNA activated protein kinase R (PKR) (Gale et al., 1998). Together with core, NS5A has been reported to associate with LDs (Brass et al., 2002, Shi et al., 2002), specifically with Apolipoprotein E (ApoE), potentially important for viral assembly and release of infectious viral particles (Benga et al., 2010).

**NS5B protein**

NS5B is an RNA-dependent RNA polymerase (RdRp), the catalytic subunit of the HCV replicase complex. The catalytic subunit consists of glycine - aspartate - aspartate (GDD) motif that produces catalytic activity. The C-terminal α-helical transmembrane domain is responsible for post-translational targeting to the ER membrane (Behrens et al., 1996, Brass et al., 2002, Egger et al., 2002, Hwang et al., 1997, Wolk et al., 2000). NS5B can form oligomers in vitro and may catalyse RNA synthesis (Qin et al., 2002, Wang et al., 2002). The crystal structure of NS5B shows a typical polymerase structure resembling a right hand with palm, finger and thumb subdomain. The palm domain contains the fully encircled active site. Finger and thumb interact with each other to create a tunnel through which a ssRNA is directed to the active site (Ago et al., 1999, Bressanelli et al., 1999, Lesburg et al., 1999).
NS5B has also been shown in interact with NS3 through the protease domain and to cause up to a five-fold increase in helicase activity (Zhang et al., 2005).

1.2.5 Translation

In contrast to cellular capped mRNAs, which are translated via cap-dependent mechanism, the uncapped RNA molecules of viruses, such as flaviviruses, are under the control of the IRES and translated in a cap-independent IRES-mediated process (Tsukiyama Kohara et al., 1992, Wang et al., 1993). The IRES mediates internal initiation of translation through binding of the 40S ribosomal subunit to form a stable pre-initiation complex, without the requirement for canonical translation initiation factors (Otto et al., 2002, Spahn et al., 2001). Cellular protein eukaryotic initiation factor-3 (eIF-3) is then recruited binding to stem-loop III and associates with the 40S. A ternary complex is formed between eIF-2, GTP and the initiator tRNA forming eIF-2: GTP: Met-tRNA<sub>i</sub>Met, which positions the initiator tRNA at the P site of the 40S subunit (Ji et al., 2004, Lukavsky et al., 2000, Otto and Puglisi, 2004). eIF5 enhances the recognition of the start codon by the eIF-2: GTP: Met-tRNA<sub>i</sub>Met complex and acts as a GTPase-activator for eIF2 and promotes the release of the initiator tRNA. A second GTP hydrolysis step involving eIF5B enables the 60S ribosomal subunit to associate, forming the functional 80S ribosome that initiates viral protein synthesis (Ji et al., 2004, Otto and Puglisi, 2004, Sizova et al., 1998, Kieft et al., 2001). The three distinct translation initiation complexes (40S, 48S and 80S) generated, have been shown by in vitro translation experiments in HeLa S10 cells and rabbit reticulocyte lysates and by ex vivo experiments in mammalian cells (Kong and Sarnow, 2002).

1.2.6 RNA replication

RNA replication occurs at the membranous web via the generation of a minus strand RNA intermediate using the positive-strand RNA as a template (Lohmann et al., 1999). This was first identified in the human hepatoma cell line Huh7 containing the
sub-genomic HCV replicons (Gosert et al., 2003). As mentioned above the membranous web is induced by NS4B and is probably ER derived (Egger et al., 2002). The second step involves the negative strand RNA serving as a template to produce multiple strands of positive polarity that will be used for polyprotein translation, packaged into new viral particles or used for synthesis of new replication intermediates (Bartenschlager et al., 2004). NS5B RdRp was thought to catalyse RNA synthesis through primer-dependent initiation, through elongation of a primer hybridized to the RNA template or through a copy-back mechanism (Behrens et al., 1996). HCV RdRp has more recently been shown to initiate de novo RNA synthesis under certain experimental conditions (Zhong et al., 2000). An additional host factor involved in RNA replication is CypB, which was shown to interact with NS5B stimulating its RNA binding activity (Watashi et al., 2005, Heck et al., 2009).

1.2.7 Transmission routes associated with hepacivirus and pegivirus infection

HCV and HPgV are most efficiently transmitted through repeated or large percutaneous exposure to infected blood, through blood transfusions, transplantations, and injecting drug use. Heat inactivation of plasma products from 1985 led to a considerable decline in infection rates with enveloped viruses. Transmission of HCV prior to 1990 was largely associated with blood transfusions and the receipt of non-heat-inactivated blood products, such as factor VIII and IX. Subsequently the development of the first-generation HCV blood test made it possible to diagnose infection in potential donors (Alter and Houghton, 2000). A study examining transfusion-associated HCV infection in haemophiliacs found that 80% of the patients were HCV positive (Jarvis et al., 1996b). As a consequence of routine and effective HCV screening in the developed world transfusion-associated HCV infections have now been virtually eliminated (Schreiber et al., 1996), but it remains an active source of HCV transmission in poorer developing countries, where blood screening is not routine (Hladik et al., 2006).
A second contributory factor for the transmission of HCV has been unsafe therapeutic injections performed by both professionals and non-professionals. The WHO estimated in 2000 that 2 million new HCV infections are acquired annually from contaminated health care injections, accounting for up to 40% of all HCV infections worldwide (Hauri et al., 2004). Between the 1960s and 1980s the mass anti-schistosomiasis campaigns in Egypt and the use of unsterilized needles and syringes by authorities has been directly linked to high levels of HCV infection in the country (Darwish et al., 1993, Frank et al., 2000, El-Zayadi et al., 1997, Darwish et al., 1996).

In more recent times the predominant mode of transmission in the developed world has been injecting drug use and the frequent use and re-use of contaminated needles between users. It now accounts for most newly acquired infections in the United States, in addition to Northern, Western and Southern Europe. Although infection rates among injecting drug users (IDUs) in developed countries have declined dramatically from 80% to between 10-30% since the late 1980s (Hope et al., 2001, Des Jarlais et al., 2003), they remain extremely high (up to 70%) in countries of the second world, such as Bulgaria and Vietnam (Vassilev et al., 2006, Quan et al., 2009).

A study on the efficiency of vertical transmission of HCV found an overall rate of just 2.7%, restricted to infants born to viraemic mothers, but this rate increased to 5.4% in women co-infected with HIV (Ferrero et al., 2003, Lauer and Walker, 2001). HCV is also far less efficiently transmitted by mucosal exposures to blood or serum-derived fluids (e.g., sex with an infected partner) or by single small dose percutaneous exposures (e.g., accidental needle sticks). Occupational transmissions of HCV infection through contaminated needle stick injuries are generally confined to health care workers with an average incidence rate of 1.8% (Puro et al., 1995a, Puro et al., 1995b, Yazdanpanah et al., 2005).
Sexual transmission of HCV is inefficient but possible, but highly dependent upon sexual behaviour. Individuals in long-term monogamous relationships are less likely to be infected (0-0.6% per year) than those who regularly change their partners (0.4-1.8% per year). The observed differences may be due to differences in sexual practices or differences in exposure rates to nonsexual sources of HCV such as the use of shared toothbrushes and shared razors (Terrault, 2002).

Other procedures potentially responsible for maintaining HCV infection through the years are cosmetic procedures such as tattooing and piercings, and religious and cultural practices such as circumcision and intranasal drug use. In developing countries with low sanitary standards, poor sterilisation of medical equipment, poor health practices and transmission though cosmetic or religious procedures is more likely to occur as discussed above in the case of anti-schistosomal therapy in Egypt (Frank et al., 2000).

Parenteral, sexual and vertical transmission have all been documented for HPgV. Prevalence studies suggest up to 4% of healthy blood donors in the developed world are viraemic at the time of donation, and a further another 5–13% have anti-E2 antibodies (Blair et al., 1998, Gutierrez et al., 1997, Pilot-Matias et al., 1996a, Tacke et al., 1997a). Rates of up to 20% have been observed amongst blood donors in some developing countries (Polgreen et al., 2003, Mohr and Stapleton, 2009). A study of 72 haemophiliacs detected HPgV infection in 14% of patients who received non-virus inactivated blood concentrates factor VIII and IX (Jarvis et al., 1996a). HPgV infection is far greater in individuals with blood-borne or sexually transmitted infections (Scallan et al., 1998), and in one study of HIV-infected homosexual men, 39.6% had viraemia and 46% had E2 antibody detected, a total exposure rate of 85.6% (Williams et al., 2004). HPgV is also efficiently transmitted vertically from mother to infant (Feucht et al., 1996, Zanetti et al., 1998) and horizontal transmission has also been documented in leprous Japanese patients (Egawa et al., 1996).
Evidence of HPgV sexual transmission has also been reported in HIV co-infected individuals, where HPgV infection was detected in 18% of co-infections (Fiordalisi et al., 1997). Additionally reports have detected high levels of HPgV viraemia amongst prostitutes (18-21%) (Kao et al., 1997a, Scallan et al., 1998a, Wu et al., 1997), and homosexual and heterosexual couples (Kao et al., 1997b, Scallan et al., 1998a). As observed with HCV, HPgV RNA has been detected in up to 40% of IDUs in the United States and anti-E2 antibodies in 70% of cases (Stapleton, 2003, Boodram et al., 2011). Higher levels of anti-E2 antibodies of up 85% have been detected in IDUs in West African specimens (Dille et al., 1997).

The recent detection of novel hepaciviruses and pegiviruses in rodents, and bats may have implications for the studies on transmission routes of these viruses. For example the hunting and consumption of bushmeat in Africa and other countries, as previously reported (LeBreton et al., 2006, Mickleburgh et al., 2009) may turn out to have historically played a major role in the origin of human viruses and may provide active reservoirs for future zoonotic transmission between bats and rodents and humans.

1.3 History and Origins of Hepatitis C Virus

Globally it is estimated approximately 170 million people are infected with HCV (WHO, 1998, Shepard et al., 2005). The aetiological agent of HCV was discovered in 1989 (clone 5-1-1) and subsequently other clones were constructed from overlapping fragments of cDNA clones (Choo et al., 1991, Takamizawa et al., 1991, Okamoto et al., 1991, Kato et al., 1990). Sequence comparison between these strains showed only 80-90% homology, reflecting the diverse nature of an RNA virus and the existence of several subtypes. From comparative analyses of these genomes, the structure and organisation of the virus was determined. It was demonstrated that the virus has an RNA genome of approximately 9,500 nucleotides (nt) with a single open
reading frame (ORF) encoding for a 3,010 to 3,033 amino acid long single polyprotein. The ORF is flanked by a well-conserved 5’UTR, postulated to be 341 nt long, and a more variable 3’UTR. The polyprotein itself could be processed by cellular proteases and viral proteases as proposed for other flaviviruses. This generated the structural proteins core (C), E1 and E2 (envelope proteins) and p7, as well as the down-stream encoded non-structural proteins NS2, NS3, NS4A, NS4B, NS5A and NS5B.

Comparative sequence analysis of the HCV genome and encoded polyprotein showed that HCV has a basic genetic organisation and polyprotein structure similar to those of pestiviruses and flaviviruses (Choo et al., 1991). Computer analysis revealed that HCV possessed statistically significant similarity to 2 sequences in the protein data bank, the NS3 protein of dengue type 2 and the putative replicase of carnation mottle virus (CARMV), a member of the Carmovirus family. This was unexpected as CARMV is a plant virus. It was suggested that this finding adds support to the hypothesis that there is an evolutionary relationship between animal and plant viruses (Miller and Purcell, 1990).

1.3.1 Genetic diversity and genotype classification

HCV shows considerable genetic diversity; 7 genotypes with >30% nt sequence divergence from each other (Simmonds et al., 2005). Until recently, GBV-B was the only other virus classified as a member of the Hepacivirus genus, recovered from a laboratory-infected tamarind (Simons et al., 1995b, Muerhoff et al., 1995). To date only a single isolate of GBV-B has been identified and the ultimate origin remains unidentified.

Of the 7 HCV genotypes several are associated with suspected endemic source areas in central and western sub-Saharan Africa (genotypes 1, 2, and 4) (Ruggieri et al., 1996, Jeannel et al., 1998, Candotti et al., 2003, Ndjomou et al., 2003) and South
East Asia (genotypes 3 and 6) (Smith et al., 1997, Tokita et al., 1994a). Additionally, extensive neighbouring regions of the African continent are dominated by individual genotypes. HCV infections in Western Africa are predominately genotype 2 (Candotti et al., 2003, Jeannel et al., 1998, Mellor et al., 1995, Ruggieri et al., 1996), while those in Central African countries of Congo, Cameroon and Gabon are attributed to genotypes 1 and 4 (Bukh and Miller, 1994, Fretz et al., 1995, Menendez et al., 1999, Xu et al., 1994, Li et al., 2009, Li et al., 2012), and limited available studies indicate genotypes 5 and 7 are largely the source of HCV infections in Central and Southern Africa (Murphy et al., 2007, Smuts and Kannemeyer, 1995).

These regions harbour the greatest diversity of HCV subtypes implying a long-term, endemic circulation of the virus over several hundred years. The spread of certain variants from these populations, such as 1a and 1b to Western countries, 3a among injection drug users in Europe, and 4a to Egypt, where it was extensively transmitted by medical injections (Pybus et al., 2003, Frank et al., 2000, Magiorkinis et al., 2009, Pybus et al., 2005), show several parallels with the emergence and rapid spread of HIV-1 among new risk groups from a central African reservoir over a similar time frame (Gao et al., 1999).

1.3.2 Dissemination of HCV worldwide

Developing countries and the Western world saw the dramatic spread of HCV during the twentieth century, through parenteral routes of transmission associated with blood transfusions, vaccinations, other medical procedures and more recently injecting drug use. The discovery of HCV in the late 1980s was driven by the concerns of clinicians and epidemiologists who increasing witnessed cases of non-A, non-B hepatitis associated with blood transfusions and medical procedures using plasma-derived blood products and unsterilized needles (Prince et al., 1974, Feinstone et al., 1975, Alter et al., 1975). Development of effective diagnostic methods made evident the extent of the viral dissemination globally. Currently an
estimated 170 million are chronically infected globally, representing 3% of the world’s population, resulting in an extensive healthcare burden from chronic liver disease and cirrhosis (Pawlotsky, 2003, Hoofnagle, 2002, Seeff, 2002). Modern transmission routes associated with injecting drug use and needle sharing is the primary route of viral transmission following the introduction in the 1990s of effective blood donor screening programs and steps to inactivate blood products (Nelson et al., 2011).

The hypothesis for the relatively recent dissemination of HCV into the western world can only be indirectly inferred, due to a lack of sample availability pre World War II. Modelling based on the evolutionary history of currently circulating variants of the virus is the best available method for indirectly inferring timescale for viral spread. HCV transmission is inefficient and rare through sexual contact or from mother to child (Wasley and Alter, 2000, Pradat and Trepo, 2000, Thomas, 2000), therefore a timescale for the spread of HCV can be inferred by examining development and availability of primary parenteral transmission routes. The use of unscreened blood in transfusions, unsterilized needles for vaccinations and in modern times, exposure to contaminated needles associated with IDU points to a recent timeframe for the spread of HCV. Such practices were not commonplace prior to World War II supporting the hypothesis for the spread of genotypes 1b and type 2 subtypes during the 1940s-50s and in the 1960s the transmission among IDUs (Pybus et al., 2001, Cochrane et al., 2002).

1.3.3 HCV geographical distribution

As previously described, HCV can be divided into 7 major genotypes which have different geographical distributions, (Simmonds et al., 2005). Amongst European blood donors and HCV patients in the 90s the predominant circulating genotypes in order of their prevalence were 1b, 2 and 3 (Touzet et al., 2000). HCV subtypes are epidemiologically distinct, targeting different risk groups and geographical regions
Introduction

that reflect their recent epidemic spread. Genotypes 3a and 1a largely infect IDUs in Northern Europe; while 4a is frequently detected in the Middle East and particularly Egypt, further evidence for the recent dissemination of HCV into these populations. (Van Asten et al., 2004, Kamel et al., 1994, Abdel-Aziz et al., 2000, Ray et al., 2000). One hypothesis for the 10-20% infection rates observed in Egypt is that extensive nationwide transmission occurred after a countrywide public health campaign inoculating against Schistosoma infections (Arthur et al., 1995, Frank et al., 2000). The effect of such measures is evident in the sharp decline in HCV infections in persons born after the discontinuation of the campaign in the 1970s (Frank et al., 2000, El-Zayadi et al., 1997, Angelico et al., 1997, Pybus et al., 2003, Kamel et al., 1994). The Egypt case study provides further supportive evidence for the recent emergence of HCV into developing countries through the utilisation of modern transmission routes.

As mentioned previously the introduction of HCV screening changed the predominant mode of transmission to IDUs and with that the genotype distribution from genotypes 1b and 2 to genotypes 1a, 3a and 4a. In Northern Europe the predominant genotypes are 1a and 3a followed by 2, while in Southern Europe genotypes 1a, 3a and 3 prevail. With the exception of Norway and the United Kingdom genotype 4 is equally distributed throughout Europe (Esteban et al., 2008). In the United States genotypes 1a and 1b are detected most frequently and to a lesser extent 2a, 2b, 3a and 4a (Shiboski and Padian, 1996). Genotype 5a was once believed to be restricted to South Africa, although recent reports suggest a high prevalence of genotype 5a in isolated regions of Central France and West Belgium for a long time (Henquell et al., 2004). Genotype 1b, 2a and 2b are associated with the most infections in Japan and other Far Eastern countries (Hara et al., 1996, Yun et al., 2008).
Highly divergent subtypes of HCV circulate in Africa, India, and South East Asia. Genotype 2 predominates in West Africa (Jeannel et al., 1998, Ruggieri et al., 1996, Candotti et al., 2003, Mellor et al., 1995), whilst genotype 1 and 4 are detected in Central African countries such as the Democratic Republic of Congo, Cameroon and Gabon (Fretz et al., 1995, Xu et al., 1994, Ndjomou et al., 2003, Bukh et al., 1993, Mellor et al., 1995, Stuyver et al., 1993, Li et al., 2012). Based on currently available findings Genotype 7 appears to be confined to Central and Southern Africa (Murphy et al., 2007). Comparable levels of genetic diversity can be observed within genotypes 3 and 6 circulating in Southern and Eastern Asia, in Vietnam, Nepal and Indonesia (Tokita et al., 1996, Tokita et al., 1994a, Tokita et al., 1994b).

As a consequence of these observations it is hypothesized that HCV has been present in the human population for an extended period of time in parts of Africa and Asia where the highest level of genetic diversity is observed, compared to industrialised countries where HCV is less diverse and therefore may have been introduced into the population more recently (Simmonds, 2001a, Simmonds et al., 2005). The availability of effective transmission routes such as blood transfusions, the use of unsterilized medical equipment and unsafe procedures, and in more recent times the sharing of needles between IDUs and the unsafe sexual behaviour among HIV infected men has resulted in the rapid dissemination of HCV from regions of endemic infection into countries of the developed world (Cochrane et al., 2002, Van deLaar et al., 2009, Pybus et al., 2001).

1.3.4 Homologs of HCV in other mammals

Based on the finding that human HIV-1 originated from chimpanzees (Gao et al., 1999.), NHP species have frequently been speculated to be the ultimate animal origin for HCV. In the wild several sub-species of chimpanzees have been found to be infected with a lentivirus derived from SIV’s infecting Old Wold monkey species (Bailes et al., 2003). Several published (Makuwa et al., 2003, Makuwa et al., 2006)
and unpublished surveys (S. Lyons et al., Chapter 5) examined the hypotheses for an equivalent NHP origin for HCV and homologous viruses in chimpanzees, apes and other Old World monkey species. These surveys were supported by the detection of a virus distantly related to HCV in a laboratory-inoculated tamarin (New World monkey), termed GBV-B (Simons et al., 1995a, Muerhoff et al., 1995). It was thought that this virus represented the New World monkey homolog of HCV and therefore hypothesised that further HCV-like viruses would be distributed among Old World monkey species in Africa and Asia. However, GBV-B infections have not been detected in any further tamarins or other New World monkey species in the wild or in captivity, and its origins like HCV, have not yet been determined. All published surveys and results presented in Chapter 5 and 6 found no PCR or serological evidence of active or past HCV or HCV-like infections in any apes or Old World monkey species (Makuwa et al., 2003, Makuwa et al., 2006).

Considering the focus on NHPs as the ultimate origin of HCV, the discovery of a canine virus (canine hepacivirus [CHV]) much more closely related to HCV than GBV-B, was unexpected and surprising (Kapoor et al., 2011). Given the host and the suspected association of the virus with infectious respiratory disease it suggests some major difference to what might be expected for HCV homologs. Subsequently CHV RNA was detected in 8 of 103 horse plasma samples in New York (Burbelo et al., 2012b). Serological testing of horses for antibodies to the conserved NS3 viral antigen determined an overall seropositivity of 35%, while 80 canine samples were PCR and antibody negative. The virus was termed non-primate hepacivirus (NPHV) as a result of the wider host range implied by the findings and this nomenclature is used this thesis. Applying sequence knowledge from the novel CHV and NPHV variants in addition to deep sequencing technology, homologs of HCV have subsequently been detected in bats (bat hepacivirus-BHV) (Quan et al., 2013) and rodents (rodent hepacivirus-RHV) (Drexler et al., 2013a, Kapoor et al., 2013b) and
most recently in Old World monkeys, black and white colobus in Uganda (GHV) (Lauck et al., 2013).

1.3.5 Disease association of hepacivirus infections

Diagnosis

Hepacivirus and pegivirus infections affect approximately 3% and 5% of the world’s population respectively (Sulaiman et al., 1995, Shepard et al., 2005, Alter et al., 1999, Desenclos, 2000, Frank et al., 2000). HCV is a leading cause of chronic liver disease, hepatocellular carcinoma and liver cirrhosis, in contrast to HPgV, which is a lymphotropic virus of unknown pathogenesis in humans and infects an estimated (Berg et al., 1999). Essential in the treatment of HCV is correct and effective diagnosis, leading to the provision of the most efficient treatment regimens (Bornschlegel et al., 2013). Currently diagnosis and screening for on-going and historical HCV infection is based on the combined detection of anti-HCV antibodies and a confirmatory PCR to detect HCV RNA (Bornschlegel et al., 2013). Detection of serological markers of HCV infection currently applies assays targeting the highly immunogenic regions of the virus core, NS3 and NS5 (Alter, 1994, Courouce et al., 1995, Courouce et al., 1994, Vallari et al., 1992, Pawlotsky et al., 1994). Diagnosis of acute HCV is difficult, primarily due to the fact that 70% of patients do not exhibit symptoms (CDC, 1998, Glynn et al., 2005). Between 10-25% of individuals with acute HCV infections will present with jaundice, nausea and vomiting. Symptoms if present usually appear 5-12 weeks post exposure, and last for 2-12 weeks. Diagnosis of acute HCV at this point is possible through the detection of rising aminotransferase levels (ALT) peaking between 400-1000IU/L. Serum bilirubin may also be elevated to levels exceeding 12mg/dL (Glynn et al., 2005).

Recently novel serology assays have been applied to other mammalian other species to detect HCV homologs, CHV and NPHV in dogs and horses respectively (Burbelo et al., 2012b). The novel LIPS assay uses antigens expressed in cells as fused Renilla
Luciferase (Ruc)-antigen recombinants and is recognised for its sensitivity and ease of production (Burbelo et al., 2009, Burbelo et al., 2012a, Ramanathan et al., 2008, Burbelo et al., 2005). An indirect immunoflorescent assay (IFA) was also recently applied using HCV-infected HuH7-cells (strain JC1) or replicon JFH1-transfected cells to screen rodent and bat samples for anti-HCV antibodies (Drexler et al., 2013a).

**Acute hepatitis**

Acute hepatitis is generally defined as the initial 6 months of infection after the presumed period of exposure during which clinical signs or symptoms of hepatitis are present (Blackard et al., 2008). The acute phase is largely asymptomatic; therefore our knowledge of the course of acute infection is limited as cases often go unreported (Berman et al., 1979, Orland et al., 2001). Serum HCV RNA levels can be detected 1-3 weeks post-infection and after a 5-12 week incubation period symptoms may develop (Farci et al., 1991). If symptoms do present, they are generally very mild and frequently mistaken for the symptoms of a common cold or flu; including nausea, vomiting, fatigue, decreased appetite and jaundice (Santantonio et al., 2003, Blackard et al., 2008). After 4-12 weeks the first increases in ALT levels indicate the first signs of liver injury (Heathcote et al., 2003). Acute HCV infection is generally diagnosed when HCV RNA is detected with anti-HCV antibodies 4-12 weeks post infection (Mondelli et al., 2005, Pawlotsky, 2002). Acute HCV is confirmed when all other causes of liver damage are excluded, when potential recent exposure is confirmed and when ALT levels reach 10-20 times the normal limit and increasing levels antibodies are detected (Mondelli et al., 2005). As antibodies are not detected in the initial 4-12 weeks RT-PCR methods are applied to detect HCV RNA in blood.

**Spontaneous resolution of acute hepatitis C**

On average 26% of all patients are able to spontaneously resolve an acute hepatitis
and clear HCV infection. It is typically most likely to happen in females and within 3 months after the onset of disease (Micallef et al., 2006). A state of chronic infection is likely to develop if the virus is not cleared within 6 months. Other studies have observed a delayed clearance between 6-12 months post infection (Larghi et al., 2002, Jauncey et al., 2004). The clinical course of the disease can be influenced by many different factors, including HCV genotype, co-infection with HIV, human leukocyte antigens (HLAs), C/C polymorphism on the IL28b gene of European and African patients, gender, race and advanced age (Thomas et al., 2009, Kenny-Walsh, 1999, Schnuriger et al., 2009). One survey in a German prison found that Caucasian men infected with genotype 3 were more likely to spontaneously clear infection than those infected with genotype 1 (Lehmann et al., 2004). A strong and multispecific cellular immune response has also been reported to play a strong role in viral clearance (Diepolder et al., 1995, Gerlach et al., 1999, Gruner et al., 2000, Thimme et al., 2001, Lucas et al., 2007).

**Chronic hepatitis**

Approximately 70-80% of infected individuals will not clear acute infection and will progress to a persistent chronic infection. The course of infection varies between patients, but is usually asymptomatic. As a consequence of HCV infection over the course of 20-30 years (up to 50 years) the virus may induce liver fibrosis, and cirrhosis and ultimately cause hepatocellular carcinoma (HCC). HCC is a major health burden in the developed world and approximately 80% of cases are the result of HBV or HCV infection (Thomas and Zhu, 2005). Inflammatory lesions on the liver are characteristic of a chronic infection in addition to the development of steatosis, the accumulation of intrahepatic lipids (Moradpour and Blum, 2005). Alcohol consumption, co-infection with HIV/HBV, obesity, age and metabolic factors may all effect the progression of the disease towards liver fibrosis and HCC (Jamal et al., 2005, El-Serag et al., 2004, Chen et al., 2008, Niederau et al., 1998, Zoulim et al., 2003, Pradat et al., 2007).
The exact mechanisms responsible for the development of HCC in chronic HCV infection remain undetermined (McGivern and Lemon, 2009). In addition to the role of HCV in HCC, fibrosis, and steatosis has also been shown to up-regulate mitogenesis with direct oncogenic potential (Koike, 2007). This oncogenic potential for hepatocyte transformation has been reported for core, E1/E2, NS2, NS3 protease, NS4A, NS4B and NS5A (Bartosch et al., 2009). An interaction between the HCV core protein and the critical tumour suppressor protein p53 has also been reported (Lu et al., 1999).

1.3.6 Prevalence of hepaciviruses in mammalian species

Globally the prevalence of HCV infection is estimated to be between 1-3%, representing 123-170 million people (WHO, 1999, Shepard et al., 2005, WHO, 1998). Epidemiological surveys worldwide largely rely upon seroprevalence studies carried out on select populations, e.g. blood donors and IDUs, which are not completely representative of the global population. Population-based studies are more informative but are not always possible in many parts of the world. Although HCV is globally distributed as previously discussed, the highest prevalence is observed in Africa and Asia, while North America, Europe and Australia are the countries of the developed world with the lowest recorded infection rates. However HCV prevalence varies considerably between countries with less than 1.0% in Northern Europe to over 2.9% in Northern Africa and Asia (Shepard et al., 2005). The lowest prevalence has been reported in Scandinavian countries and the United Kingdom (0.01-0.1%), while the highest prevalence has been documented in Egypt (10-22%), a country with an estimated population of 73 million (Bird et al., 2001b, Bird et al., 2001a, Frank et al., 2000). The Egyptian case study observed that the prevalence of infection increased with age among all age groups (Abdel-Aziz et al., 2000) and that genotype 4 was the predominant genotype infecting 90% of those studied, implying a recent shared source of exposure (Ray et al., 2000). As
previously mentioned the high incidence of HCV and genotype 4 in Egypt has been positively linked with an anti-schistosomal therapy campaign between 1960-1980, where unsterilized medical equipment was used (Frank et al., 2000).

Relatively low rates of HCV seroprevalence have been detected in Australia (1.1%) (Law, 1999), France (1.1%) (Desenclos, 2000), and Germany (0.6%) (Palitzsch et al., 1999). Slightly higher rates have been reported in the United States (1.8%) (Alter et al., 1999) and Japan (0.1-2%) (Chung et al., 2010). In developed countries it is reported that incidence rates peaked around 2000 and are in decline (Chung et al., 2010). High levels of infection have been reported in China (3.2%) a country with one fifth of the world’s population (Xia et al., 1996), and in India (0.9%) also home to another fifth of the world’s population (Chowdhury et al., 2003). A single study of blood donors in Indonesia reported incidence rate of 2.1%, while a more thorough study in Pakistan reported rates between 2.4-6.5% (Waheed et al., 2009, Luby et al., 1997, Sultana et al., 2000, Khattak et al., 2002).

Levels of viraemia have been estimated for BHV, RHV, GHV and NPHV at between 0.6-7.8%, although the total exposure rates have not been estimated amongst these species by serology based methods. NPHV infection rates of 3-7.8% were reported among horse populations in New York State and the United Kingdom (Burbelo et al., 2012b, Lyons et al., 2012a) and antibodies were further detected in 35% of horses in the US (Burbelo et al., 2012b). In a single large global study RHV was detected 1.8% of 1,465 European bank voles (Myodes glareolus) and 1.9% of 518 South African four-striped mice out of a total 4,470 rodent samples (Drexler et al., 2013a). Rates of infection among bats have been preliminary estimated at 0.6% of bats (Quan et al., 2013).
1.4 History of human pegivirus

Following the discovery of hepatitis C virus (HCV) in 1989, it became clear that HCV was not detected in 10-20% of individuals with non-A, non-B hepatitis and therefore research groups searched for a novel etiological agents responsible for non-A, non-B, non-C hepatitis (Alter, 1989, Choo et al., 1989, Kuo et al., 1989, Simons et al., 1995a). Two independent research groups working at Gene Labs and Abbott Laboratories reported the detection of a human virus with significant homology to HCV detected in patients with non-A, non-B, and non-C hepatitis in 1995 and 1996 (Linnen et al., 1996, Simons et al., 1995b). Despite the lack of epidemiological data to support an association with hepatitis Gene Labs called this virus “Hepatitis G virus” (HGV) (Linnen et al., 1996), following the detection of a putative hepatitis F virus (Deka et al., 1994). Abbott laboratories isolated two viruses from two marmosets inoculated with serum from a surgeon with non-A, non-B hepatitis with the initials G.B., and termed these isolates GBV-A and GBV-B (Schlauder et al., 1995b, Simons et al., 1995b). Degenerate primers designed by Abbott and based on the sequences of GBV-A and GBV-B were subsequently applied to detect a closely related human homolog GBV-C (Simons et al., 1995a). Subsequent sequence analysis of HGV and GBV-C revealed that both isolates were the same virus and termed HGV/GBV-C and based on phylogenetic analysis GB viruses were classified as members of the Flaviviridae family (Kim and Fry, 1997, Leary et al., 1996b).

Further research by Abbott laboratories detected GBV-C_trog in an HCV-infected chimpanzee with resolving hepatitis but it was not detected in human or macaque monkey samples (Birkenmeyer et al., 1998). Adams et al. also identified GBV-C RNA in 3 non-captive chimpanzees (subspecies troglodytes and verus) that they called GBV-C_cpz (Adams et al., 1998). Based on phylogenetic analysis the GBV-C_cpz was considered a chimpanzee variant of GBV-C rather than a separate genotype (Adams et al., 1998). GBV-B is most closely related to HCV, while GBV-A and
GBV-C/HGV form a separate cluster (Muerhoff et al., 1995, Simons et al., 1995b). A virus more distantly related to GBV-C, GBV-A and GBV-A was also detected in serum from Old world frugivorous bats and termed GBV-D (Epstein et al., 2010).

**GB Virus re-classification and novel mammalian homologs**

A recent review of the GB-viruses proposed that the nomenclature and classification of these viruses be adjusted to reflect their origins and tropism (Stapleton et al., 2011). This now accepted proposal was based on three points. Firstly, that there is no evidence that the surgeon ‘GB’ was infected with GBV-A, GBV-C or GBV-D. Secondly the fact that GBV-D appears to be confined to bats, and GBV A and GBV-B confined to New World primates means it is extremely unlikely the surgeon ‘GB’ was the origin. Finally and importantly there is no evidence that GBV-C causes hepatitis in humans. As a result of the review the viruses have been re-named and re-classified as either members of the *Hepacivirus* or *Pegivirus* virus genus (Pegivirus: ‘pe’ representing persistent, ‘g’ representing GB or G) (Table 1.4). Subsequently highly divergent pegivirus isolates have been detected in rodents (RPgV) and horses (EPgV) (Table 1.4) in the United States and United Kingdom (Kapoor et al., 2013a, Kapoor et al., 2013b).

**Table 1.4: Current nomenclature, classification and pathogenesis of original ‘GB’ viruses and hepaciviruses.** The table includes the proposed nomenclature for novel homologs identified in rodents, horses, bats and Old World monkeys.

<table>
<thead>
<tr>
<th>Current Nomenclature</th>
<th>Original Virus</th>
<th>Tropism</th>
<th>Host Range</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Pegivirus Genus</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HPgV</td>
<td>GBV-C</td>
<td>Lymphocytes</td>
<td>Humans/Chimpanzees</td>
</tr>
<tr>
<td>SPgVcpz</td>
<td>GBV-Ccpz</td>
<td>ID³</td>
<td>Chimpanzee</td>
</tr>
<tr>
<td>SPgV</td>
<td>GBV-A</td>
<td>Lymphocytes</td>
<td>NWP</td>
</tr>
<tr>
<td>BPgV</td>
<td>GBV-D</td>
<td>ID</td>
<td>Bats</td>
</tr>
<tr>
<td>EPgV</td>
<td>NA²</td>
<td>ID</td>
<td>Horses</td>
</tr>
<tr>
<td>RPgV</td>
<td>NA</td>
<td>ID</td>
<td>Rodents</td>
</tr>
<tr>
<td>TDAV</td>
<td>NA</td>
<td>ID</td>
<td>Horses</td>
</tr>
</tbody>
</table>
1.4.1 Genetic diversity of members of the Pegivirus genus

Pegiviruses infect a diverse range of mammals including humans [HPgV] (Simons et al., 1995a, Simons et al., 1995b, Yoshiha et al., 1995), bats [BPgV](Epstein et al., 2010), new world monkeys and chimpanzees [SPgV/SPgV_{cpz}] (Adams et al., 1998, Birkenmeyer et al., 1998, Bukh and Apgar, 1997, Epstein et al., 2010, Kapoor et al., 2013a, Leary et al., 1996a, Linnen et al., 1996, Muerhoff et al., 1995, Quan et al., 2013, Simons et al., 1995b) rodents [RPgV] and horses [EPgV and TDAV] (Kapoor et al., 2013a, Kapoor et al., 2013b, Quan et al., 2013). HPgV variants are currently assigned to 6 genotypes (Muerhoff et al., 2006, Muerhoff et al., 2005) and an additional genotype 7 has been proposed for isolates recovered from IDUs in Yunnan, China (Feng et al., 2011). Genotypes exhibit a strong geographical association; genotype 1 predominating in Africa (Liu et al., 2003), genotype 2 in Europe (Brancio et al., 2010), genotype 3 in Asia, genotype 4 in Southeast Asia, genotype 5 in South Africa and genotype 6 in Indonesia (Muerhoff et al., 2006). The global distribution of genotypes follows a distinct pattern reflecting the migration routes of humans out of Africa, suggesting the HPgV has co-evolved with humans (Pavesi, 2001, Smith et al., 2000). Despite this perceived ancient origin, compared to HCV (>30%) HPgV genetic diversity is lower with between 11-14% nucleotide divergence between genotypes (Muerhoff et al., 2006, Pavesi, 2001). Variants isolated from chimpanzees (SPgV/SPgV_{cpz}) are significantly more diverse than HPgV, further supporting the hypothesis of co-evolution (Birkenmeyer et al., 1998, Adams et al., 1998). The SPgV_{cpz} genome shares 84% amino acid identity with

<table>
<thead>
<tr>
<th>Pegivirus Genus</th>
<th>HPgV</th>
<th>SPgV/SPgV_{cpz}</th>
<th>RPgV</th>
<th>BPgV</th>
<th>EPgV and TDAV</th>
</tr>
</thead>
<tbody>
<tr>
<td>HCV</td>
<td>HCV</td>
<td>Hepatocytes</td>
<td>Humans/Chimpanzees</td>
<td>GBV</td>
<td>GBV-B Hepatocytes</td>
</tr>
<tr>
<td>GBV</td>
<td>NA</td>
<td>ID</td>
<td>Horses</td>
<td>NPHV</td>
<td>NA</td>
</tr>
<tr>
<td>RHV</td>
<td>NA</td>
<td>ID</td>
<td>Rodents</td>
<td>RVH</td>
<td>NA</td>
</tr>
<tr>
<td>BHV</td>
<td>NA</td>
<td>ID</td>
<td>Bats</td>
<td>GHV</td>
<td>NA</td>
</tr>
<tr>
<td>ID, insufficient data</td>
<td>NA; not applicable</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

1. ID, insufficient data
2. NA; not applicable
SPgV, while HPgV isolates were >95% identical (Mohr and Stapleton, 2009, Muerhoff et al., 2005, Pavesi, 2001). Based on alignment with HPgV and HCV the SPgVcpz genome is predicted to have the same genome organisation and protein functions (Birkenmeyer et al., 1998).

Active pegivirus infections documented in horses [EPgV] display 62-77% sequence divergence from other known pegivirus sequences in the structural genes E1, E2 and X, and between 49-59% in non-structural regions NS2-NS5 (Kapoor et al., 2013a). Little or no homolog was observed in the NS4A and NS5A regions and comparable to other members of the pegivirus genus the most highly conserved regions were the NS3 helicase and NS5B RNA polymerase (Kapoor et al., 2013a).

Global studies of wild rodent and bat populations have detected rodent [RPgV] and bat pegivirus [BPgV] infections and have revealed a much greater viral diversity in members of the Pegivirus genus (Kapoor et al., 2011, Kapoor et al., 2013a, Burbelo et al., 2012b, Lyons et al., 2012a, Quan et al., 2013, Drexler et al., 2013a, Lauck et al., 2013). BPgV isolates were discovered widely distributed across continents, in the New and Old world and phylogenetic analysis suggests that BPgV may be the ultimate reservoir for these viruses (Quan et al., 2013, Epstein et al., 2010). Regions of highest sequence conservation were observed in NS3 and NS5B of BPgV with 43-79% and 39-71% amino acid sequence identity respectively (Quan et al., 2013). RPgV has also been shown to display high genetic diversity, with isolates displaying 78-81% amino acid divergence to HPgV, BPgV, SPgV and EPgV in structural regions and 54-56% in non-structural regions (Kapoor et al., 2013a). Certain sub-species of bats have been shown to be natural reservoirs for many zoonotic viruses that cause disease in humans, including paramyxoviruses, lysaviruses, and severe acute respiratory syndrome (SARS)-related coronaviruses (Luis et al., 2013, Calisher et al., 2006, Lau et al., 2005, Drexler et al., 2012, Chua et al., 2000, Li et al., 2005). Bats also possess certain characteristics that may contribute to their ability to act as
natural reservoirs for viruses, their long lifespan of up to 35 years, their high populations densities, their habitat and roosting behaviour, high levels of species diversity and unique immune systems where particular viruses cause persistent infections but bats remain healthy, while homologs of these viruses are pathogenic to humans, and other mammals (Calisher et al., 2006).

1.4.2 Prevalence of pegivirus infections in mammalian species

HPgV infection is globally distributed and capable of establishing a persistent infection (Barnes et al., 2007, Lefrere et al., 1997, Lefrere et al., 1996). HPgV infection is common, and it is estimated that between 1- 5% of healthy blood donors in developed countries are viremic at the time of donation. The level of viraemia in blood donors of developing countries is considerably higher approaching 20% in some countries (Blair et al., 1998, Gutierrez et al., 1997, Pilot-Matias et al., 1996a, Tacke et al., 1997a). The detected prevalence is higher in blood donors from developing countries and significantly higher in patients with co-infected with blood-borne or sexually transmitted infections like HCV and HIV (Alter et al., 1997a, Stapleton, 2003). Among individuals infected with HCV and HIV the HPgV prevalence rates approach 20% and 30% respectively (Alter et al., 1997a, Stapleton, 2003). A single study of HIV infected homosexual men detected HPgV viraemia in 39.6% and antibodies in 46%, a combined infection rate of ~86% (Williams et al., 2004). In contrast to HCV, HPgV antibodies are not usually detected during viraemia, and antibodies to the HPgV E2 protein develop after viral clearance (Pilot-Matias et al., 1996a, Surowy et al., 1997, Tacke et al., 1997b, Tanaka et al., 1998a, Thomas et al., 1998, Tacke et al., 1997a). E2 antibodies are thought to contain neutralizing activity, as anti-E2 antibodies appear to provide protection against re-infection (Hassoba et al., 1998) (Tillmann et al., 1998, Thomas et al., 1998). RT-PCR is used to detect active viraemia in sera or plasma and detection of past infection relies on the detection of anti-E2 antibodies (Alter, 1997, Stapleton, 2003,
Stapleton et al., 2004). Prevalence studies may underestimate the rates of past exposure as E2 antibodies may disappear during longitudinal follow up (Schwarze-Zander et al., 2006, Stark et al., 1996).

Preliminary studies in horses, rodents and bats have observed frequencies estimated at between 3.9-5% (Quan et al., 2013, Kapoor et al., 2013a, Kapoor et al., 2013b). However there is currently no available data to estimate the total level of exposure among these mammalian species worldwide and no virus-specific serology based assays have been designed to determine the rate of past infection.

1.4.3 Disease associations and diagnosis of pegivirus infections

Diagnosis

Infection with HPgV, in contrast to HCV is not associated with hepatic disease and interest in the virus waned when studies failed to detect any disease associations (Batts, 1997). Diagnosis of acute and chronic HPgV infection is based on the detection of anti-E2 antibodies in the presence or absence or viral RNA (Tacke et al., 1997a, Tacke et al., 1997b, Dille et al., 1997). Nucleic acid testing (NAT) and the application of next generation sequencing technologies has led to the detection of pegivirus infection in addition to those infecting humans and NHPs in a range of mammals including rodents (RPgV), bats (BPgV) and horses (EPgV) (Quan et al., 2013, Epstein et al., 2010, Simons et al., 1995a, Kapoor et al., 2013a, Kapoor et al., 2013b). Currently serological based screening like that applied in determining HPgV seroprevalence has not been possible in bats, rodents and horses but the determination of the level of infection amongst these species and the possible detection of more divergent viral species using antibody detection assays will contribute to our understanding of pegivirus and hepacivirus infection, pathogenesis, host immune responses and the elusive origins of HCV and HPgV in humans (Simmonds, 2001a, Pavesi, 2001, Tanaka et al., 1998c).
Tropism and pathogenesis

HPgV and SPgV/SPgV<sub>cpz</sub> are readily detected in the circulating lymphocytes of infected hosts and at low to non-detectable levels in the liver, suggesting that these viruses could be lymphotropic, and not hepatotropic (Kobayashi <i>et al.</i>, 1999, Laskus <i>et al.</i>, 1997, Pessoa <i>et al.</i>, 1998, Radkowski <i>et al.</i>, 1999, Radkowski <i>et al.</i>, 2000, Simons <i>et al.</i>, 2000, Tucker <i>et al.</i>, 2000). Pegiviruses originally termed ‘GB’ viruses were thought to be associated with hepatic disease, although the reclassification of HPgV (GBV-C), BPgV (GBV-D), SPgV (GBV-A), EPgV and RPgV as members of the <i>Pegivirus</i> genus is a reflection of a lack of evidence suggesting an association with hepatitis (Stapleton <i>et al.</i>, 2011).

HPgV RNA is detected and produced in T and B-lymphocytes based on infected patients studied <i>ex vivo</i> (George <i>et al.</i>, 2006, Mellor <i>et al.</i>, 1998). Primary human peripheral blood mononuclear cells (PBMCs) are the most widely used cell culture system for <i>in vitro</i> growth of HPgV, implying that HPgV is a lymphotropic virus (Fogeda <i>et al.</i>, 1999, Fogeda <i>et al.</i>, 2000, Xiang <i>et al.</i>, 2000). However questions still remain about the primary site of HPgV replication and negative strand RNA is either very low or undetectable in PBMCs of infected individuals (Mellor <i>et al.</i>, 1998) but in one study was detected in bone marrow and spleen samples (Laskus <i>et al.</i>, 1998). HPgV infection, as previously discussed, is transmitted by blood-borne, vertical, and sexual routes, consistent with lymphotropism (Bourlet <i>et al.</i>, 1999, Fiordalisi <i>et al.</i>, 1997, Berzsenyi <i>et al.</i>, 2005, Bjorkman <i>et al.</i>, 1998). The clinical significance of infection with BPgV (GBV-D), SPgV<sub>cpz</sub> (GBV-C<sub>cpz</sub>), SPgV (GBV-A), EPgV and RPgV has not been determined. Current findings predominately suggest no association between hepatitis and HPgV infection (Theodore and Lemon, 1997, Feucht <i>et al.</i>, 1997). BPgV has been detected from clinically healthy bats at concentrations between $10^3$-$10^8$ RNA copies per ml of plasma, although levels in the liver and other cells have not been analysed (Quan <i>et al.</i>, 2013, Epstein <i>et al.</i>, 2010). Viral loads of between $10^{4.5}$ and $10^{6.5}$ were detectable in the plasma of horses infected
with EPgV, and RNA was further detected in the liver, lymph and PBMCs of two chronically infected horses but there was notifiable difference in RNA levels between the tissues (Kapoor et al., 2013a).

Theiler’s disease is an acute serum hepatitis in horses most often associated with the administration of the tetanus antitoxin. Until recently the aetiology responsible for the disease was unknown but an infectious agents transported through blood transfusions was suggested (Hjerpe, 1964, Pearson, 1999). Some infections occur in horses who have not received equine serum products, but who have been exposed to animals that have received such products (Tennant, 1978). Clinical symptoms usually occur 4-10 weeks after exposure, and while cases can be sporadic, several epidemics have also been reported (Guglick et al., 1995). Infection with the recently isolated TDAV infection contrasts with other members of the Pegiviurs genus as findings from three separate outbreaks of acute serum hepatitis suggest an association with liver disease in horses (Chandriani et al., 2013). The tissue tropism of pegiviruses infecting horses remains to be conclusively determined through the study of active infection.

**HPgV-HIV Interaction**

When studies on HPgV pathogenesis in HIV infected patients indicated an inhibitory effect on HIV replication interest was revived (Devereux et al., 1998, Hollingsworth et al., 1998, Mohr and Stapleton, 2009, Tillmann et al., 2001). HPgV research was initially performed by viral hepatitis research groups, and the realization that HPgV did not cause hepatitis resulted in a marked reduction in research activity. The observation that HBV and HCV co-infection alters the clinical course and outcome of HIV-infected patients, led researchers to also hypothesise that HPgV may also impact the progression of HIV infection among co-infected individuals. In 1998 two groups studied the level of HPgV viraemia and seroprevalence and correlated the results with clinical follow-up data, and noticed that HIV-infected individuals co-
infected with HPgV survived longer than those without HPgV (Heringlake et al., 1998, Toyoda et al., 1998). These preliminary results were confirmed by several subsequent studies, although not all (Lefrere et al., 1999, Van der Bij et al., 2005, Williams et al., 2004, Xiang et al., 2001, Yeo et al., 2000, Tillmann et al., 2001). It was also evident that active HPgV viraemia was necessary to for this association as patients observed to clear viraemia have a worse prognosis (Van der Bij et al., 2005, Williams et al., 2004, Tillmann et al., 2001). HPgV viraemia is associated with an improved clinical response to antiretroviral therapy (ART) (Souza et al., 2006, Rodriguez et al., 2003). HIV-infected patients treated with anti-retroviral therapies and co-infected with HPgV also have significantly higher survival rates compared to HPgV negative patients (Tillmann et al., 2001).

1.5 Other points of consideration

1.5.1 Diagnostic methods applied in detection of equine hepatopathy

Liver disease is relatively common occurrence in horses and foals suffering metabolic, septic, hypoxic and neoplastic conditions. Progression to liver failure when approximately 75% of the liver tissue is damaged is fortunately rare. Early liver fibrosis can be reversed and treated promptly when effectively diagnosed. While histological evaluation of the liver tissue is diagnostically the gold standard in equine hepatic disease diagnosis, biopsies are an invasive procedure. Diagnosis of hepatic disease can be achieved through non-invasive measures such as biochemical analysis of serum and ultrasound scans.

Early signs of liver insufficiencies include depression and a decreased appetite, although these symptoms may not be observed. Horses may develop more evident symptoms such as jaundice (yellowing visible in the eyes, gums, and skin), severe weight loss, anorexia, abdominal pain and colic, photosensitisation, personality and behavioural changes and less often, digestive problems such as diarrhoea, and
Introduction

constipation (Divers, 2005, McGorum *et al.*, 1999, Pearson, 1990, Divers *et al.*, 1988, Pearson, 1999). A single study of 50 horses presenting with indicators of liver insufficiencies found that ill demeanour (34 cases), decreased appetite (28 cases), colic (25 cases), and weight loss (25 cases) were some of the most common clinical signs of hepatic disease (McGorum *et al.*, 1999). These initial presenting symptoms are the early indicators to veterinarians of potential underlying liver disease and insufficiencies.

Biochemical analysis of liver enzymes is an imperative next step in the diagnosis, and the design of the most efficient treatment method, improving the ultimate prognosis before liver failure occurs (Smith *et al.*, 2003). Abnormal liver enzymes occur when 60-70% of the liver function is lost in the horse and will likely indicate increased levels of bilirubin, ammonia, and serum iron and gamma globulins in the case of chronic disease. Therefore it is important that when studying liver disease in horses to observe a combination of liver enzymes and to apply careful interpretation of the results (Durham *et al.*, 2003a, Durham *et al.*, 2003b, Smith *et al.*, 2003, McGorum *et al.*, 1999). Sorbital dehydrogenase (SDH) and gamma glutamyltransferase (GGT) are liver specific enzymes that respectively can reflect hepatocellular and biliary injury. In addition aspartate aminotransferase (AST) and alkaline phosphatase (ALP) can indicate hepatocellular and biliary injury, but are not strictly liver specific. Damage to hepatocytes causes AST to elevate, although elevated AST may also be an indicator of muscle damage. Alanine aminotransferase (ALT) often used in the diagnosis of liver disease in small animals does not leak at significant levels from equine hepatocytes so this enzyme is not a useful diagnostic tool for hepatic disease in horses.

Although mild to moderate increases in serum GGT (e.g. up to 140 IU/L) are of limited diagnostic or prognostic value, it is nevertheless very unusual to find significant hepatopathy in horses in the absence of increased serum GGT (McGorum
et al., 1999, Durham et al., 2003a, Durham et al., 2003b). GGT is considered the most sensitive test for liver disease in the horse (Divers, 1993, West, 1989b, Divers et al., 1988). A single study reported sensitivity of up to 75% and specificity of 90% when using GGT as a primary screening test for subclinical liver disease in horses exposed to pyrrolizidine alkaloids, while ALP had a substantially lower specificity of 58% (Curran et al., 1996). Signs of liver disease can be observed in horses with pyrrolizidine alkaloid poisoning. Liver enzymes become elevated before clinical signs of disease appear (Lessard et al., 1986, Mendel et al., 1988), but GLDH returns to normal or below normal levels before clinical signs are detected. As a consequence, serum GGT, GLDH and total bile acid concentrations are more sensitive indicators of this chronic condition. Elevation of serum concentrations of GGT can be detected after a few days of liver damage and remain elevated until the terminal phase (Craig et al., 1991). GGT values >400 IU/L are associated with a poor prognosis. Elevated GGT can frequently be observed in racehorses (50-140 IU/L) due to lesions of the biliary system or drug-induced increases in GGT (Divers, 2005). For the purpose of all studies presented within this thesis GGT levels >42U/L were considered as indicative of underlying hepatic insufficiencies, although only when parallel enzyme elevations were observed in one or both GLDH and bile acid.

Elevations in GLDH are frequently observed in the presence of acute hepatocellular damage (McGorum et al., 1999). It is a mitochondrial enzyme mainly associated with the liver where it is found at highest concentrations but is also found in cardiac muscle and the kidneys. GLDH is ideal for diagnosis of acute hepatocellular damage in combination with other enzymes as GLDH levels can normalise during chronic liver disease (Craig et al., 1991). The sensitivity of GDH in the diagnosis of hepatic disease is reported to be 63%, exceeding that of SDH (Durham et al., 2003a). GLDH tests in combination with GGT and bile acids are indicators of hepatocellular damage and hepatic function respectively. GLDH levels >12U/L combined with elevated GGT and bile acids were considered as indicative of liver insufficiencies for the
purpose of this thesis. Serum bile acid levels when measured in combination with GLDH are useful indicators of hepatocellular damage (West, 1989a). Increased serum bile acids are indicative of bile flow obstructions and hepatocellular damage (Pearson and Craig, 1992). The main limitation in measuring serum bile acid is that liver disease must be severe before significant elevations are observed (McGorum et al., 1999) and liver disease cases can be found to display normal bile acid levels but with marked increases in other enzymes for example GLDH and GGT. Analysis of the physical presenting symptoms is an important aid in the interpretation of bile acids as a lack of appetite and anorexia can increase serum bile acids as high as 20-30 µmol/L in the absence of liver disease. Horses with elevations in serum bile acids greater than 20 µmol/L are less likely to survive than those with lower values and chronic cases with concentrations greater than 100 mmol/L are almost invariably fatal (McGorum et al., 1999).

Liver biopsies are the definitive method of diagnosis, although they are not always possible or desirable in horses (Durham et al., 2003a, Durham et al., 2003b, Smith et al., 2003). Biopsies can provide significant information regarding the underlying disease, the extent of liver damage and potentially the cause of the liver insufficiencies in the animal.

1.5.2 Causative agents associated with equine hepatic disease

The susceptibility of the liver to disease is based on its role in the clearance of toxins and drugs from the body, and as a consequence grazing animals like horses are particularly prone to hepatic disease (West, 1996). Equine hepatic disease is common in horses and foals suffering from septic, hypoxic, neoplastic, or metabolic conditions. A range of aetiological causes may be ascribed to the development of liver disease, including toxins, infectious agents, viruses, metabolic diseases, and obstructions of the liver and digestive system (Divers, 2005)(Table 1. 5).
The most common cause of chronic liver disease in horses in the United States is caused by the ingestion of plants that contain pyrrolizidine alkaloids (PA) found in ragwort (Senecio jacobae), tar weed (Senecio trianularis), and rattlebox (Crotalaria spp.) among others (Divers, 1983, Knight et al., 1984). For induction of liver disease the horses needs to consume 2-5% of its body weight (Stegelmeier et al., 1996).

Table 1.5: Aetiological agents of equine hepatic insufficiencies, their sources and the clinical indicators resulting. (Adapted from Table 2 (Pearson, 1999))

<table>
<thead>
<tr>
<th>Substance</th>
<th>Source</th>
<th>Pathology or clinical sign</th>
</tr>
</thead>
<tbody>
<tr>
<td>Toxic: Pyrrolizidine alkaloids</td>
<td>Compound found in pasture plants: e.g. Ragwort, tar weed, Rattlebox</td>
<td>Necrosis, fibrosis, cirrhosis, weight loss.</td>
</tr>
<tr>
<td>Panicum grass (Panicum spp.)</td>
<td>Kleingrass with toxin sapogenin</td>
<td>Biliary fibrosis, necrosis</td>
</tr>
<tr>
<td>Alsike Clover (Difolium hybridum)</td>
<td>Ingestion of clover pasture or hay</td>
<td>Portal fibrosis, biliary hyperplasia</td>
</tr>
<tr>
<td>Iron</td>
<td>Feed supplements, leaching into water or feed.</td>
<td>Portal necrosis, &gt;iron content</td>
</tr>
<tr>
<td>Mycotoxins</td>
<td>Ingestion mouldy feed</td>
<td>Centralobular necrosis, biliary hyperplasia</td>
</tr>
<tr>
<td>Infectious: Cholangiohepatitis</td>
<td>Salmonella sp.</td>
<td>Bile duct proliferation, colon damage</td>
</tr>
<tr>
<td>Tyzzer’s disease</td>
<td>Clostridium piliforme</td>
<td>Hepatitis, jaundice, fever, diarrhoea</td>
</tr>
<tr>
<td>Parasites</td>
<td>Parascaris equorum</td>
<td>Necrosis, fibrosis, lymphocytic inflammation, hepatic amyloidosis</td>
</tr>
<tr>
<td>Non-Infectious</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hepatic lipidosis</td>
<td>Lipid metabolism disorder</td>
<td>Decrease appetite, weakness, lethargy</td>
</tr>
<tr>
<td>Neoplasia</td>
<td>Carcinomas</td>
<td>Melanomas, lesions, papillomas, equine sarcoids</td>
</tr>
<tr>
<td>Chronic active hepatitis</td>
<td>Infection, immune response, toxins</td>
<td>Jaundice, behaviour and appetite change, weight loss, depression</td>
</tr>
<tr>
<td>Granulomatous disease</td>
<td>Unknown, linked to schistosomiasis, exaggerated immune response, allergens</td>
<td>Weight loss, digestive problems, colic, depression, leukocytosis</td>
</tr>
</tbody>
</table>

Obstructive
Introduction

<table>
<thead>
<tr>
<th>Biliary stones</th>
<th>Hyperbilirubinemia, Altered enzyme levels, fever, weight loss</th>
</tr>
</thead>
<tbody>
<tr>
<td>Portal vein thrombosis</td>
<td>Acute is asymptomatic, cirrhosis</td>
</tr>
<tr>
<td>Hepatic torsion</td>
<td>Colic, blood count abnormalities</td>
</tr>
<tr>
<td>Liver Lobe Atrophy</td>
<td>High-concentrate, low-fibre diets, long-term compression</td>
</tr>
<tr>
<td>Colic, digestive disorders</td>
<td></td>
</tr>
</tbody>
</table>

**Unknown Agents**

<table>
<thead>
<tr>
<th>Disease</th>
<th>Serum/Blood, Virus TDAV</th>
<th>Mood change, weight loss, liver enzyme elevation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Theiler’s Disease (potential viral origin)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Other toxic agents associated with equine hepatic disease include panicum grass (Johnson *et al.*, 2006), alsike clover (Knight, 1995), (Nation, 1991), iron (Lewis, 1995) and mycotoxins such as Aflatoxin B1 and B2 (Pier *et al.*, 1980, Vesonder *et al.*, 1991). Infectious agents attributed to hepatic insufficiencies include parasites and Tyzzer’s disease. Tyzzer’s disease is attributed to infection with *Clostridium piliforme* spreading from the lower intestine via blood and lymph (Kahn, 2006) (Table 1.5). The most significant parasites infecting the equine liver are *P. equorum* and *F. hepatica* (Kornsas *et al.*, 2006, Epe *et al.*, 2004). Cholangiohepatitis is associated with *Salmonella* sp., *E. coli*, *Pseudomonas* sp. or *Actinobacillus equuli* leading to bile duct and liver inflammation (Kahn, 2006). Chronic active hepatitis is often related to cholangiohepatitis, and is a histological diagnosis based on the observation in the liver of a sustained inflammatory process (Kahn, 2006).

Cholelithiasis refers to the presence of gallstones in the gallbladder. Horses do not possess a gallbladder, however the presence of gallstones in the biliary tract are the most common cause of biliary duct obstructions in horses (Gerros *et al.*, 1993). The causes of these stones in horses have not been conclusively determined but changes in cholesterol levels and salt concentrations in the body have been suspected. Additionally parasites such as *Clonorchis sinensis* or *Opisthorchis viverrini* and the migration of parasite *Ascaris lumbricoides* into the common bile duct can result in bile duct occlusion (Gerros, 1996).
Hepatic neoplasia is not often observed in horses but metastasis of the liver and bile ducts is a preliminary indicator of liver carcinoma (Kahn, 2006). Hepatic torsion, neoplasia, right lobe compression, and thrombosis are obstructive causes of liver insufficiencies. Granulomatous disease is rare and of unknown aetiology, that can form cutaneous lesions and tumour like nodules on the liver as well as other organs (Scott and Miller, 2003).

Acute equine serum hepatitis has been found to be associated with the receipt of blood and serum products. It is the most common cause of acute hepatitis. Recent findings suggest that a virus termed Theiler’s Disease Associated Virus is the aetiologial agent responsible for the disease (Chandriani et al., 2013, Aleman et al., 2005). The detection of novel equine pathogens EPGV a homolog of HPgV associated with persistent infections in humans and NPHV a homolog of HCV in humans causing acute and chronic hepatitis in humans, suggests that as yet undiscovered viruses may be associated with equine hepatic diseases (Lyons et al., 2012a, Kapoor et al., 2013a, Burbelo et al., 2012b).

1.6 Aims of this thesis

Despite the substantial economical and health burdens associated with major human pathogens such as HBV and HCV, their ultimate origins have remained elusive and poorly understood (WHO, 1998, WHO, 1999, Simmonds, 2001a, Simmonds, 2001b). The absence of such information pertaining to the evolutionary history of HBV, HCV and their genetically related viruses impacts upon the development of vaccines and effective eradication strategies. Studies are currently limited by the absence of historical samples from which to date the emergence of human infections and therefore the evolution of human hepatic viruses relies on epidemiological studies and genetic analysis of contemporary virus populations worldwide.
Introduction

Approximately one third of the world’s population is infected with HBV, and despite the availability of a vaccine, the virus is attributed with over 1 million deaths per year through liver disease. The past 15 years has seen the detection of HBV in chimpanzees, gorillas and other non-human primates (NHPs) at frequencies comparable to those observed in regions of endemic human HBV infection (MacDonald et al., 2000, Takahashi et al., 2000). Despite the genetic divergence between human and NHP HBV variants the detection of recombination between human genotype C and chimpanzee and gibbon variants suggests that HBV can share hosts in nature. Our main ambitions in screening NHP samples from Cameroon, a region of endemic HBV infection, was to determine the genotypes infecting each species, the presence if any of HBV recombinants in NHPs and the potential cross-species transfer between geographically-proximal species e.g. gorillas and chimpanzees. The detection of a recombinant gorilla/chimpanzee HBV variant and the observation that chimpanzee sub-species specific variants were transmitted between species allowed us to hypothesise an evolutionary process that may explain the genetic diversity, geographical distribution and species-specificity observed for HBV. Results presented in chapter 3 suggest that allopatric speciation may account for these observations and the generation of novel NHP HBV recombinants is possible between species that are geographically proximal in the wild (Lyons et al., 2012b).

HCV and HPgV infect approximately 3-5% of the world’s population and HCV is one of the major causes of chronic liver disease, hepatocellular carcinoma and liver cirrhosis. At the inception of this project we sought to screen NHPs from a region of endemic HCV infection with the ultimate goal of detecting a NHP homolog. However based on the unexpected detection of HCV homologs in canines (CHV) (Kapoor et al., 2011) and horses (NPHV) (Burbelo et al., 2012b), the research was broadened to the screening of a diverse range of mammalian species including dogs, horses, cats, and donkeys. Large scale screening of these species, as documented in chapter 4 resulted in the detection and clinical characterisation of NPHV infection among horses in the UK (Lyons et al., 2012a) and
Introduction

presented preliminary novel data to suggest that unlike HCV, in horses NPHV infection is non-hepatotrophic but may establish a persistent infection with extremely high viral loads at the peak. Based on these findings in horses we additionally screened by PCR for the presence of HPgV homologs. Data generated in this prevalence study is presented in chapter 5 and comparable with HPgV, findings imply that EPgV is capable of establishing a persistent infection, which is not hepatotrophic and is of unknown aetiology (Lyons et al., 2014 Submitted).

Seroprevalence data from a single US study detected NPHV antibodies in 35% of horse samples (Burbelo et al., 2012b). To expand upon this initial study, methods outlined in chapter 6 were applied to generate two novel ELISA assays to detect NPHV (NS3 and core) and EPgV (NS3) antibodies. Applying these assays to all the mammalian serum and plasma available we concluded that both NPHV and EPgV are endemic in the horse population, with 43% and 63% seroprevalence respectively. Furthermore the detection of NPHV antibodies among stable-mates and a single dog living with a viraemic horse supports the initial detection of CHV RNA in dogs in kennels in US, suggesting the potential for zoonotic transmission of the virus between species (Kapoor et al., 2011) and also implies that transmission may be associated with close contact between animals or exposure to contaminated blood or other bodily fluids. Research presented within this chapter extends data currently available on the seroprevalence and infection frequencies of NPHV and EPgV.

Through examination of serial EPgV RNA positive samples (Chapter 5) the relationship between antibody detection and the presence or absence of viral RNA was examined over the course of infection and used as a potential marker to indicate the presence or absence of persistent infection in the host. The detection of antibodies in horses across a wide age range including yearlings, suggests that the routes of transmission used by HPgV, parenteral, vertical and sexual may also be responsible for the high EPgV seroprevalence detected.
Chapter 2
Materials and Methods

Materials and methods used in completing this thesis comprise both laboratory techniques and computational sequence analysis. Laboratory methods were applied firstly to the screening of a large collection of samples for the presence of novel viruses homologous to human pathogens, secondly to generating sequence fragments from positive isolates for evolutionary analysis, and finally to the generation of recombinant viral proteins for the development of serology assays, described in Section 2.2. Computational and bioinformatics techniques were then employed to generate phylogenetic and recombination analyses, determination of substitution rates and structural prediction of genomic regions like 5’UTR. These are described in Section 2.3. Archives of samples from the Royal (Dick) School of Veterinary Studies and Metabiota (Formerly Global Viral Forecasting Initiative [GVFI]) were central to all aspects of methodology applied in this thesis as described in Section 2.1.

2.1 Storage and archiving of clinical samples

Since 2006, large archives of clinical samples submitted for virological testing and virus discovery, have been maintained by both the Royal (Dick) School of Veterinary Studies and Professor P. Simmonds in conjunction with Metabiota. Samples include a range of species and sample types as detailed in each chapter. All non-human primate Metabiota samples were referred for virological testing to detect viruses homologous to human HCV, HPgV and HBV. All excess diagnostic companion animal samples from R(D)SVS were referred for determination of CHV, NPHV and EPgV exposure. Viral diagnostics are an essential component for predicting prognosis, especially where effective treatments are not always available for a large number of viruses that are routinely screened for in humans and companion animals,
and allow for the generation of disease surveillance maps and efficient protocols for prevention of infection.

### 2.2 Laboratory techniques

#### 2.2.1 Prevention of contamination of polymerase chain reaction products

Results generated in this thesis were dependent on a large amount of sequence data generated by polymerase chain reaction (PCR) amplification of NPHV, EPgV and HBV from positive samples followed by Sanger sequencing. PCR is a highly sensitive diagnostic technique and is potentially subject to contamination by extraneous DNA sequences at a number of critical stages and minimisation and exclusion of such contaminants is critical for all studies discussed in this thesis. PCR contamination can severely compromise the evolutionary analysis of amplified genomes and the detection of specific virus in large numbers of samples.

Good laboratory practice and aseptic techniques were employed at all stages in the PCR process with precautions to ensure the exclusion of contaminants. Additionally, all primers and reagents used in the preparation of PCR master mixes were kept in small aliquots specific for each individual in the laboratory. DNA was added to the PCR reaction mixture after all other reagents in order to reduce possibility of false positives (Kwok and Higuchi, 1989). No template controls prepared from the same master mix were also included to detect any PCR reagent contamination.

Separate designated areas and equipment were used to reduce the potential of carryover contamination. Carryover occurs when the reaction mixtures are contaminated with products from previous reactions. In order to limit the effect of carryover a one-way laboratory system was implemented for each stage of PCR, a process that involved the unidirectional transfer of reagents, samples, amplification products and equipment between and within certain areas (Figure 2.1).
The preparation of buffers, primers, nucleotides and reagent mixes was carried out in a designated clean room containing specific equipment and a strict one-way system with no re-entry of removed reagents, personal protective equipment, or containers. Two areas were reserved specifically for sample preparation including, defrosting, extraction, nucleic acid synthesis and cataloguing, containing area-specific personal protective equipment (PPE), reagents, equipment and freezer storage. Where RNA transcripts of known high copy number were to be included in the PCR (Lyons et al., 2012a, McLeish et al., 2012) and the contamination risk associated with such controls was high, serial dilutions were prepared in segregated areas distinct from those of extraction and PCR reagent preparation.

In addition to the clean-room, laboratory areas were designated specifically for cDNA and first round PCR products (Figure 2.1), second round PCR and sequencing reactions, and finally gel electrophoresis (Figure 2.1). All primers used for sequencing reactions were prepared in the clean-room and adhered to the one-way system. By complying with all these protocols group members have ensured that the occurrence of PCR contamination during the course of experiments for this thesis has been minimal.
FIGURE 2.1: Unidirectional laboratory system for prevention of PCR contamination in reactions. Arrows denote steps between areas that are designed to flow in one direction. Transport of samples, reagents, or equipment backwards was not permitted in these areas. White arrow denotes the movement from building 1 to building 2. Red area indicates all unidirectional procedures occurring in designated areas of building 1. Green area is designated gel electrophoresis only. Owing to space restrictions second round PCR preparation and sequencing reactions were carried out in the same laboratory area (marked with blue box). Every effort was still made to designate physical bench space, reagents and pipettes as second round PCR and sequencing only.

2.2.2 Extraction of RNA from clinical samples

RNA extractions were performed on 140ul of plasma or respiratory samples using the QI Amp Viral RNA extraction kit (Qiagen) according to the manufacturer’s instructions and eluted in final volume of 60ul. 140 µl of plasma was mixed with 560µl of Buffer AVL containing 5.6 µl Buffer AVE-carrier RNA (1µg/µL carrier RNA) and mixed by pulse vortexing for 15seconds followed by incubation at room temperature for 10minutes and brief centrifugation. 560 µl of 96-100% ethanol was added to the sample, mixed by pulse vortexing for 15seconds and briefly centrifuged to recover droplets from the lid. 630ul of solution was added to QI Amp Mini Column and centrifuged at 7656 xg for 1min and the filtrate is discarded. This step is repeated until all lysate has been loaded on to the spin column. RNA bound to the silica membrane was washed with 500ul of AW1 Buffer and centrifuged at 7656 xg for
1 min. A full speed centrifuge at 12000 xg for 3 minutes followed a further wash with 500ul of AW2, centrifuged. The column was placed in an RNase free 1.5ml eppendorf and 60ul of room temperature Buffer AVE was applied to the silica membrane and allowed to incubate at room temperature for one minute. To elute the RNA, the column was spun at 10,000 xg for one minute. RNA was stored at -20°C for immediate use and -40°C for long term.

2.2.3 Isolation of PBMCs from whole blood

HISTOPAQUE®-1077 is a non-sterile reagent used for the isolation of mononuclear cells from whole blood. HISTOPAQUE®-1077 is a polysucrose and sodium diatrizoate solution adjusted to a density of 1.077± 0.001 g/ml. Anticoagulated venous blood is layered onto histopaque solution in centrifuge tube. Through centrifugation aggregates of granulocytes and erythrocytes rapidly sediment; while lymphocytes and other monocytes remain at the interface between the plasma and the HISTOPAQUE®-1077 solution. Contamination of PBMCs with erythrocytes is negligible and extraneous platelets can be subsequently removed by centrifugation at low speeds during the washing steps.
FIGURE 2.2: Histopaque 1077 isolation of peripheral blood mononuclear cells from whole blood. Ficoll-histopaque is placed at the bottom of a centrifuge conical tube (Step 1), and an equal volume of whole blood is carefully layered on top. Following centrifugation (Step 2), the following layers are visible; listed from top to bottom of the centrifuge tube, a layer plasma and other constituents, a layer of mono-nuclear cells called PBMCs, followed by a layer Ficoll histopaque solution, and finally a layer of erythrocytes and granulocytes in a pellet at the bottom of the tube. This method allows for easy separation and recovery of PBMC’s from whole blood.

PBMCs were isolated as per the manufacturer’s instructions (Sigma Aldrich) from horse plasma to determine the presence of a reservoir for NPHV infection. Briefly 3.0mls of HISTOPAQUE®-1077 were added to 15ml conical centrifuge tube and allowed to equilibrate to room temperature (Figure 2.2; Step 1). An equal volume of whole blood was carefully layered on top and centrifuged at 400 x g for 30 minutes at room temperature. Following centrifugation a layer of plasma is visible on top of a layer of mononuclear cells, followed by the histopaque solution and a pellet of erythrocytes and granulocytes (Figure 2.2; Step 2). The upper layer is then aspirated to within 0.5 cm of the opaque interface containing the mononuclear cells and then discarded. The opaque interface was transferred to a clean conical and washed with 10mls of isotonic PBS solution was added and centrifuged for 250 x g for 10 minutes, mixed by gentle aspiration and centrifuged at 250 x g for 10 minutes. The
supernatant was aspirated and discarded, and the cell pellet re-suspended in 5mls of PBS and mixed by gentle aspiration and centrifuged at 250 xg for 10 minutes. The wash step is repeated and the cell pellet is re-suspended in 0.5mls of PBS. RNA was then extracted from PBMCs using QIAmp RNA blood mini kit as instructed (QIAGEN) and eluted in final volume of 100 μL.

2.2.4 Reverse Transcription

2.2.4.1 Two step Reverse transcription PCR

For the two-step RT-PCR single stranded cDNA was generated by reverse transcription (RT) allowing extracted viral RNA to be amplified by PCR. RT-PCR was performed using A3500 Reverse Transcription System (Promega, UK) according to manufacturers’ instructions in a 20 μl reaction volume which contained the following: 500 ng total or 5 μl viral RNA and DNase/RNase-free water to a final volume of 8 μl. The mixture was heated for 10 minutes at 70°C and then chilled on ice for 5 minutes. The tubes were kept on ice and 4 μl of 25mM MgCl₂, 2 μl reverse transcription 10X buffer, 2 μl of 10mM dNTPs, 0.5 μl recombinant RNasin ribonuclease inhibitor, 0.5 μl of 0.1mM random primers and 15 Units (U) of AMV reverse transcriptase were added. The reaction mixture was first incubated for 10 minutes at room temperature, then for 50 minutes at 42°C. To inactivate the reverse transcriptase, the reaction mixture was incubated for 5 minutes at 95°C. The generated cDNA was stored at -20°C.
### TABLE 2.1: Reagents included in reverse transcription reactions

<table>
<thead>
<tr>
<th>Reagent</th>
<th>Volume</th>
</tr>
</thead>
<tbody>
<tr>
<td>25mM MgCl</td>
<td>4µL</td>
</tr>
<tr>
<td>10X reverse transcription buffer¹</td>
<td>2µL</td>
</tr>
<tr>
<td>10mM deoxynucleotide triphosphates (dNTPs)</td>
<td>2µL</td>
</tr>
<tr>
<td>Recombinant RNasin ribonuclease inhibitor (25U/µL)</td>
<td>0.5µL</td>
</tr>
<tr>
<td>100µM random hexamers</td>
<td>1µL</td>
</tr>
<tr>
<td>Avian myeloblastoma virus reverse transcriptase (10U/µL)</td>
<td>0.6µL</td>
</tr>
<tr>
<td>Nuclease free H₂O</td>
<td>4.9µL</td>
</tr>
</tbody>
</table>

All positive and negative controls for use in subsequent PCR reactions were included in the RT step to ensure consistency of results. Random primers were used in the RT reaction to generate cDNA to analyse multiple genomic regions, generate complete genomes and to screen for multiple RNA viruses. Alternatively gene specific primers were used to increase the sensitivity of the RT reaction specifically targeting a virus and genomic region of interest. Gene specific techniques were applied in samples that proved difficult to amplify, although as this method precludes the use of the generated cDNA in any subsequent PCR it is less cost effective and not preferable.

#### 2.2.4.2 cDNA Synthesis by Superscript III Reverse Transcription

Screening for CHV and EPgV was carried using Superscript III reverse transcription as detailed in by the authors (Kapoor et al., 2011, Lyons et al., 2012a). Reverse transcription was carried out in a 20 µl reaction volume which contained the following: 10 µl viral RNA and 1 µl of 0.1.mM random hexamers. This solution was heated at 80°C and 95°C for 2 minutes each. On ice 4 µl of 5X buffer², 2 µl DTT, 1.25 µl 10mM dNTPs, 1 µl Superscript III reverse transcriptase and 0.75ul nuclease

---

¹ 10X reverse transcription buffer was supplied with the Reverse Transcription System kit (Promega, UK) and was composed of 100mM Tris-HCl (pH 9.0 at 25°C), 500mM KCl and 1% Triton® X-100.
² 5X Buffer contains 250 mM Tris-HCl (pH 8.3 at room temperature), 375 mM KCl, 15 mM MgCl₂
free water were added. The reaction mixture was heated at 25°C for 10 minutes, 42°C for 60 minutes and 70°C for 5 minutes. Generated cDNA was stored at -20°C.

2.2.4.3 Combined cDNA synthesis and First Round PCR

SuperScript™ III One-Step RT-PCR System with Platinum® Taq DNA polymerase was used to amplify the complete genome of NPHV and to determine the sequence of all NS3 positive NPHV and EPgV isolates. The 20µl reaction mixture was prepared on ice. Reagents used in the SSIII RT-PCR are listed in Table 2.2 and the thermal cycling carried out in Table 2.3.

<table>
<thead>
<tr>
<th>TABLE 2.2: Superscript III RT-PCR Reagents</th>
<th>TABLE 2.3: Superscript III Protocol</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reagent</td>
<td>Volume</td>
</tr>
<tr>
<td>2X Reaction Buffer&lt;sup&gt;3&lt;/sup&gt;</td>
<td>10.0 µl</td>
</tr>
<tr>
<td>Sense Primer (10OD)</td>
<td>0.1 µl</td>
</tr>
<tr>
<td>Anti-sense Primer (10OD)</td>
<td>0.1 µl</td>
</tr>
<tr>
<td>Extracted Viral RNA</td>
<td>6.0 µl</td>
</tr>
<tr>
<td>Superscript&lt;sup&gt;TM&lt;/sup&gt; III Platinum Taq&lt;sup&gt;4&lt;/sup&gt;</td>
<td>0.6 µl</td>
</tr>
<tr>
<td>DNase/RNase Free Water</td>
<td>3.2 µl</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

All reactions were prepared on ice and contained reagents as detailed (Table 2.2). RNA directly extracted from clinical specimens was used. As the outer PCR primers for the particular reaction were used for the RT step, the resulting product could only be used for amplification of that specific segment. The Platinum Taq used in the amplification is complexed with an activity-blocking antibody that renders it unreactive at room temperature. This prevents non-specific annealing and therefore increases sensitivity. The complexed antibody is denatured at the high temperatures used in the thermal cycling (Table 2.3) and polymerase activity resumes. The first

---

<sup>3</sup> 2X Reaction Mix is supplied with the Superscript III One-step RT-PCR kit and contains 4mM of each dNTP and 3.2mM of MgSO<sub>4</sub>

<sup>4</sup> In contrast to standard reverse transcription reactions, RT undertaken with the Superscript III system used Moloney murine leukaemia virus reverse transcriptase (M-MLV-RT), as opposed to AMV-RT.
round product obtained from the Superscript III reaction was used directly in second round of PCR, as previously described.

2.2.5 PCR and sequencing of NPHV, EPgV, HBV positive samples

Within the studies presented in this thesis, PCR had a dual role. Firstly, samples were screened by PCR based methods (Chapter 3) for the presence of novel viruses homologous to human pathogens and secondly, specific sequence fragments were amplified for studies of genetic diversity, recombination and evolution. All amplicons where required were sequenced to enable phylogenetic analysis to be carried out using SSE v1.1 and Mega 5.0 (Tamura et al., 2011, Simmonds, 2012).

2.2.5.1 PCR Primer Design

The design of PCR primers for use in the amplification of various genome regions to complete genomes of all of NPHV, EPgV, and HBV required a delicate balance between primer degeneracy (to counteract high genetic diversity within regions) and primer specificity (to avoid amplification of non-target sequences). All PCR primers were designed in-house, either as part of a previous study (MacDonald et al., 2000) or specifically for the studies described. Upon initial design, each primer set was tested for sensitivity using serial 10-fold dilutions of known positive clinical samples, where available.

The first step in designing PCR primers was to obtain all relevant sequences from Genbank. These aligned sequences were then inspected for areas of high sequence conservation around the target regions. Regions of high sequence conservation were then analysed and a fragment for the PCR primer was selected according to the following criteria:

- High sequence conservation – as few degenerate bases as possible

---

5 In 2012, known copy number RNA transcripts were developed for NPHV and CHV (McLeish et al., 2012). After the advent of these RNA transcripts, sensitivity tests were carried out using known copy number samples at 10-fold dilutions.
Materials and Methods

- Length of 23-25 nucleotides (not less than 16 and not more than 26)
- Melting temperature (Tm) predicted to be close to 60°C
- G+C content of 40-60% with 3’ GC clamp where possible
- No self-annealing or hairpin formation

Calculations of Tm and analysis for self-annealing/hairpin formation were carried out in with the aid of the OligoCalc program (available at http://www.basic.northwestern.edu/biotools/oligocalc.html). Primer pairs for use in the same PCR reaction were selected to have a Tm of within 5 degrees of each other, where possible. The specificity of each primer set was confirmed by analysis with primer BLAST (www.ncbi.nlm.nih.gov/tools/primer-blast) to ensure that primers did not match human or other non-target sequences in the sample.

**TABLE 2.4: Codes for individual nucleotides and degenerate bases used for PCR primer design**

<table>
<thead>
<tr>
<th>Base Code</th>
<th>Nucleotides</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Adenine</td>
</tr>
<tr>
<td>C</td>
<td>Cytosine</td>
</tr>
<tr>
<td>G</td>
<td>Guanine</td>
</tr>
<tr>
<td>T</td>
<td>Thymine</td>
</tr>
<tr>
<td>K</td>
<td>G or T</td>
</tr>
<tr>
<td>M</td>
<td>A or C</td>
</tr>
<tr>
<td>S</td>
<td>G or C</td>
</tr>
<tr>
<td>W</td>
<td>A or T</td>
</tr>
<tr>
<td>Y</td>
<td>T or C</td>
</tr>
<tr>
<td>R</td>
<td>G or A</td>
</tr>
<tr>
<td>B</td>
<td>G, C or T</td>
</tr>
<tr>
<td>V</td>
<td>G, C or A</td>
</tr>
<tr>
<td>D</td>
<td>G, A or T</td>
</tr>
<tr>
<td>H</td>
<td>A, C or T</td>
</tr>
<tr>
<td>N</td>
<td>A, C, G or T</td>
</tr>
</tbody>
</table>

**2.2.5.2 PCR protocols**

Polymerase chain reaction (PCR) was performed to amplify nucleic acids for downstream applications. All PCR and RT-PCR reactions were performed using a TECHNE Flexigene thermocycler. Depending on the downstream application, three
different thermostable polymerases were used. For diagnostic PCR and for sequencing of amplicons standard GoTaq DNA polymerase (Promega) was used. PCR reactions were set up in 50µl reaction volume. Reagents used in the GoTaq PCR are listed in Table 2.5 and cycling conditions in Table 2.6.

<table>
<thead>
<tr>
<th>TABLE 2.5: GoTaq PCR reagents</th>
<th>TABLE 2.6: GoTaq PCR protocol</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Reagent</strong></td>
<td><strong>Quantity</strong></td>
</tr>
<tr>
<td>5XGoTaq Buffer</td>
<td>10 µl</td>
</tr>
<tr>
<td>3mM dNTPs^a</td>
<td>1.0 µl</td>
</tr>
<tr>
<td>Sense Primer (100uM)</td>
<td>0.3 µl</td>
</tr>
<tr>
<td>Antisense Primer (100uM)</td>
<td>0.3 µl</td>
</tr>
<tr>
<td>Template DNA^c</td>
<td>2.0 µl</td>
</tr>
<tr>
<td>GoTaq DNA Polymerase (5U/ul)^b</td>
<td>0.2 µl</td>
</tr>
<tr>
<td>DNase/RNase-free water</td>
<td>to 50 µl</td>
</tr>
</tbody>
</table>

^a dNTPs were acquired as separate dATP, dGTP, dCTP and dTTP 25mM aliquots (Promega, UK). These were then diluted to make 3mM stocks for use in PCR.

^b Tag polymerase was GoTaq® DNA Polymerase (Promega, UK).

^c For the first round of PCR, 2µL cDNA was incorporated into the reaction mixture. For the second round of PCR, 1-2µL of first round product was used.

Conditions outline in Table 2.5 and 2.6 are standard protocols and all new primers were initially tested using these conditions. Individual PCR protocols and primer sets required optimisation prior to use on large screening samples. This involved using 10-fold dilution series of positive control samples, where available, and sequentially varying certain parameters to determine the optimal set of conditions for each PCR reaction. Frequently optimised parameters included annealing temperature and increased concentrations of dNTPs. Sensitivity of several NPHV protocols was improved by inclusion of one first round primer in the second round (hemi-nested PCR). All modifications to the standard protocol used for individual PCRs are given below (Table 2.13), along with a reference to the thesis chapter to which these reactions relate.
Fragments that could not be amplified successfully using GoTaq and which were GC rich in sequence were amplified using AccuPrime Taq DNA Polymerase High Fidelity (Life Technologies). Reactions were carried out in 25ul volume. AccuPrime reagents are detailed in Table 2.7 and cycling conditions in Table 2.8.

<table>
<thead>
<tr>
<th>TABLE 2.7: AccuPrime Reagents</th>
<th>TABLE 2.8: AccuPrime PCR Protocol</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reagent</td>
<td>Quantity</td>
</tr>
<tr>
<td>10X AccuPrime Buffer*</td>
<td>2.5 µl</td>
</tr>
<tr>
<td>Sense Primer</td>
<td>0.3 µl</td>
</tr>
<tr>
<td>Antisense Primer</td>
<td>0.3 µl</td>
</tr>
<tr>
<td>Template DNA</td>
<td>4.0 µl</td>
</tr>
<tr>
<td>AccuPrime DNA Polymerase</td>
<td>0.2 µl</td>
</tr>
<tr>
<td>DNase/RNase-free water</td>
<td>to 25 µl</td>
</tr>
</tbody>
</table>

Fragments used for cloning were generated using HotStar Polymerase Kit or HotStar HiFidelity Polymerase Kit (Qiagen), annealing conditions varied based on the primer and PCR reaction. All primers were made up to a concentration 100mM, sequences are shown in Appendix A. PCR reactions were set up in 50ul reaction volume. Reagents and cycling conditions used for Hot Star HiFi PCR and Hot Star PCR are listed in Tables 2.9, Table 2.10 and Table 2.11 and Table 2.12 respectively. A non-template control (deionised (d) H₂O) was included in every PCR run to test possible contamination of reagents. PCR products were analysed by agarose gel electrophoresis (1-2% depending on the amplicon size).

*10X AccuPrime buffer contains 600 mM Tris-SO4 (pH 8.9), 180 mM (NH₄)₂SO₄, 20 mM MgSO₄, 2 mM of each dNTP, thermostable AccuPrime™ protein, and 10% glycerol.
### Materials and Methods

#### Table 2.9: Hot Star HiFi PCR Reagents

<table>
<thead>
<tr>
<th>Reagent</th>
<th>Quantity</th>
</tr>
</thead>
<tbody>
<tr>
<td>10X HotStar Buffer</td>
<td>5.1 µl</td>
</tr>
<tr>
<td>MgCl</td>
<td>3.0 µl</td>
</tr>
<tr>
<td>dNTPs</td>
<td>1.25 µl</td>
</tr>
<tr>
<td>Sense Primer (0.1 mM/ml)</td>
<td>1.0 µl</td>
</tr>
<tr>
<td>Antisense Primer (0.1 mM/ml)</td>
<td>1.0 µl</td>
</tr>
<tr>
<td>Template DNA</td>
<td>3.0 µl</td>
</tr>
<tr>
<td>HotStar DNA polymerase (3-2.5U/reaction)</td>
<td>0.65 µl</td>
</tr>
<tr>
<td>DNase/RNase-free water</td>
<td>To 53 µl</td>
</tr>
</tbody>
</table>

#### Table 2.10: Hot Star HiFi PCR Protocol

<table>
<thead>
<tr>
<th>Cycles</th>
<th>Temperature</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>95°C</td>
<td>7 minutes</td>
</tr>
<tr>
<td>10</td>
<td>95°C</td>
<td>40 seconds</td>
</tr>
<tr>
<td></td>
<td>60°C → 56°C</td>
<td>45 seconds</td>
</tr>
<tr>
<td></td>
<td>72°C</td>
<td>60 seconds</td>
</tr>
<tr>
<td>30</td>
<td>95°C</td>
<td>35 seconds</td>
</tr>
<tr>
<td></td>
<td>54°C</td>
<td>40 seconds</td>
</tr>
<tr>
<td></td>
<td>72°C</td>
<td>60 seconds</td>
</tr>
<tr>
<td>1</td>
<td>72°C</td>
<td>10 minutes</td>
</tr>
</tbody>
</table>

#### Table 2.11: Hot Star PCR Protocol

<table>
<thead>
<tr>
<th>Reagent</th>
<th>Quantity</th>
</tr>
</thead>
<tbody>
<tr>
<td>10X HotStar HiFi Buffer</td>
<td>10 µl</td>
</tr>
<tr>
<td>Sense Primer (0.1 mM/ml)</td>
<td>0.3 µl</td>
</tr>
<tr>
<td>Antisense Primer (0.1 mM/ml)</td>
<td>0.3 µl</td>
</tr>
<tr>
<td>Template DNA</td>
<td>1 ng</td>
</tr>
<tr>
<td>HotStar DNA polymerase (2.5U/µl)</td>
<td>0.5 µl</td>
</tr>
<tr>
<td>DNase/RNase-free water</td>
<td>To 50 µl</td>
</tr>
</tbody>
</table>

#### Table 2.12: Hot Star PCR Reagents

<table>
<thead>
<tr>
<th>Cycles</th>
<th>Temperature</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>95°C</td>
<td>5 minutes</td>
</tr>
<tr>
<td>30</td>
<td>94°C</td>
<td>18 seconds</td>
</tr>
<tr>
<td></td>
<td>50°C</td>
<td>21 seconds</td>
</tr>
<tr>
<td></td>
<td>72°C</td>
<td>90 seconds</td>
</tr>
<tr>
<td>1</td>
<td>72°C</td>
<td>5 minutes</td>
</tr>
</tbody>
</table>

---

7 10X reaction buffer contains Tris-Cl, KCl, (NH4)2SO4, 7.5 mM MgSO4, bovine serum albumin, Triton X-100, Factor SB; pH 8.7 (20°C), 1.5 mM dNTPs (Ultrapure quality).

8 5X reaction buffer contained Tris-Cl, KCl, (NH4)2SO4, 7.5 mM MgSO4, bovine serum albumin, Triton X-100, Factor SB; pH 8.7 (20°C), 1.5 mM dNTPs (dATP, dCTP, dGTP, and dTTP (ultrapure).
**TABLE 2.13: Modifications to standard PCR protocols used for individual PCR reactions**

<table>
<thead>
<tr>
<th>PCR Reaction</th>
<th>Primers Set(s)</th>
<th>Modifications to standard reagents</th>
<th>Modifications to standard cycling conditions</th>
<th>Chapter Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>HBV Genome Walking</td>
<td>1. HBV S3/S6 2. HBV S3/26 3. HBV 15/42 4. HBV 2/40 5. HBV 3/28</td>
<td>None</td>
<td>• 35 cycles first round  • 40 cycles second round</td>
<td>Chapter 3</td>
</tr>
<tr>
<td>HBV Genome Walking</td>
<td>HBV 48/51→49/50 HBV 56/59→57/58</td>
<td>None</td>
<td>Touch Down PCR:  • 30 Cycles: Annealing temperature decrease 1°C/2 cycles:65°C-51°C  • Followed by 15 cycles of standard conditions</td>
<td>Chapter 3</td>
</tr>
<tr>
<td>HBV Genome Walking</td>
<td>HBV 52/55→53/54 →52/54 →55/53 Nested and Hemi-Nested PCRs above</td>
<td>None</td>
<td>Touch Down PCR:  • 30 Cycles: Annealing temperature decrease 0.5°C/cycle 74°C→60  • Followed by 15 cycles of standard conditions</td>
<td>Chapter 3</td>
</tr>
<tr>
<td>Mitochondrial Sequencing</td>
<td>PrC01(OS)/PrC01(OAS)</td>
<td>None</td>
<td>35 Cycles first and second round PCR</td>
<td>Chapter 3</td>
</tr>
<tr>
<td>Colony PCR</td>
<td>M13 Forward M13 Reverse</td>
<td>None</td>
<td>95°C- 2 minutes 30 cycles:  • 95°C - 30 seconds  • 50°C - 30 seconds  • 72°C – 30 seconds 72°C – 5minutes</td>
<td>Chapter 3/4</td>
</tr>
<tr>
<td>CHV Screening 1st Sound PCR</td>
<td>CHV XOF1/XOR1</td>
<td>Qiagen HotStar Taq Polymerase Kit – 3µl cDNA</td>
<td>• Touch Down PCR 40 Cycles (Outlined Above)</td>
<td>Chapter 4</td>
</tr>
<tr>
<td>CHV Screening 2nd Round PCR</td>
<td>CHV XOF2/XOR2</td>
<td>Qiagen HotStar Taq Polymerase Kit -1µl PCR product  5 µl 10X Hot Star Buffer</td>
<td>Touch Down PCR 40 Cycles: (Outlined Above)  • Touch down 10 cycles 62°C to</td>
<td>Chapter 4</td>
</tr>
</tbody>
</table>
**Materials and Methods**

<table>
<thead>
<tr>
<th>Step</th>
<th>Details</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Nonprimate hepacivirus NS3</strong></td>
<td>NPHV_NS3</td>
</tr>
<tr>
<td></td>
<td>• RNase/DNase free water up to 51µl</td>
</tr>
<tr>
<td></td>
<td>• 30 Cycles annealing temperature 57ºC</td>
</tr>
<tr>
<td></td>
<td>• 10mM dNTPs</td>
</tr>
<tr>
<td></td>
<td>• 2µl First round PCR added to second round</td>
</tr>
<tr>
<td></td>
<td>• 35 cycles in first and second round PCR</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>pGex-2t Screening</strong></td>
<td>pGex5'/pGex3'</td>
</tr>
<tr>
<td></td>
<td>None</td>
</tr>
<tr>
<td></td>
<td>35 Cycles:</td>
</tr>
<tr>
<td></td>
<td>• 94ºC-1 minute</td>
</tr>
<tr>
<td></td>
<td>• 55ºC-1 minute</td>
</tr>
<tr>
<td></td>
<td>• 72ºC-2 minutes</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Initial Equine pegivirus Screening</strong></td>
<td>DD-dF2/dR2</td>
</tr>
<tr>
<td></td>
<td>Qiagen Hot Star Taq Polymerase</td>
</tr>
<tr>
<td></td>
<td>PCR: 95ºC -8minutes activation</td>
</tr>
<tr>
<td></td>
<td>• 10 Cycles:</td>
</tr>
<tr>
<td></td>
<td>o 95ºC-40seconds</td>
</tr>
<tr>
<td></td>
<td>o 61ºC-45seconds</td>
</tr>
<tr>
<td></td>
<td>o 72ºC-30seconds</td>
</tr>
<tr>
<td></td>
<td>• 30 cycles:</td>
</tr>
<tr>
<td></td>
<td>o 95ºC-30seconds</td>
</tr>
<tr>
<td></td>
<td>o 57ºC-40seconds</td>
</tr>
<tr>
<td></td>
<td>o 72ºC-30seconds</td>
</tr>
<tr>
<td></td>
<td>• 72ºC- 10 minutes</td>
</tr>
<tr>
<td></td>
<td>• 68ºC and 66ºC annealing temperature used in second round PCR</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Equine pegivirus 5’UTR</strong></td>
<td>EPgV_5’UTR</td>
</tr>
<tr>
<td></td>
<td>Hemi Nested:</td>
</tr>
<tr>
<td></td>
<td>• First Round: IS and OAS.</td>
</tr>
<tr>
<td></td>
<td>• Second round IS and IAS.</td>
</tr>
<tr>
<td></td>
<td>Annealing temperature of 48ºC in the second round</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Equine pegivirus 5’UTR</strong></td>
<td>EPgV-NS5B_9222 [\rightarrow9780]</td>
</tr>
<tr>
<td></td>
<td>10mM dNTPs</td>
</tr>
<tr>
<td></td>
<td>2µl of first round PCR added to second round</td>
</tr>
<tr>
<td></td>
<td>Annealing temperature of 48ºC in the first round and 49ºC in the second round</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Equine pegivirus 5’UTR</strong></td>
<td>EPgV_5’UTR</td>
</tr>
<tr>
<td></td>
<td>2µl of first round PCR added to second round</td>
</tr>
<tr>
<td></td>
<td>Annealing temperature of 48ºC in the second round</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### Materials and Methods

<table>
<thead>
<tr>
<th>Equine pegivirus NS3</th>
<th>EPgV NS3</th>
<th>• 10mM dNTPs • 2µl first round PCR in second round</th>
<th>35 cycles in first and second round PCR</th>
<th>Chapter 6</th>
</tr>
</thead>
</table>

#### 2.2.5.3 Gel Electrophoresis

Agarose gel electrophoresis was used to analyse for the presence of DNA and RNA and to detect PCR positive samples. Depending upon the size of the product 0.7-2.0% agarose gel was prepared using 1X Tris-acetate-EDTA buffer (TAE) (Seven Biotechnologies, UK) with 0.2µl ethidium bromide (EtBr) or 0.1µl SYBR Safe per ml of agarose\(^9\). 10µL of second round PCR product was thoroughly mixed with 1.5µL of 6X loading dye and loaded directly into a well of prepared gel in an electrophoresis tank containing 1X TAE buffer. As the density of the loading dye is significantly greater than the TAE buffer, it allows the sample to sink into the well. The negatively charged dye co-migrates with the sample through the gel allowing for monitoring of the sample as it runs. An exACTGene Low Range 1kB Plus DNA ladder (Fisher Scientific, UK) was loaded alongside the PCR products.

All second round PCR products were electrophoresed at 150 volts for 40-50 minutes, depending on the size of the expected product. EtBr and SYBR safe intercalates into the DNA and exhibits fluorescence under UV light, thereby allowing the visualisation of the DNA/RNA.

#### 2.2.5.4 Gel extraction and purification of amplified products

In cases where multiple products were visualized on an agarose gel and the PCR reaction could not be further optimised, bands were excised and purified using the PureLink™ Quick Gel Extraction and PCR Purification Combo Kit (Invitrogen) according to the manufacturers’ instructions. Briefly, DNA bands were manually

---

\(^9\) Ethidium bromide was used for the majority of initial screening studies undertaken during this thesis. However, as it is an intercalating agent and is a known carcinogen with prolonged exposure, certain University areas discourage its use. For later studies, SybrSAFE was used as an alternative.
excised under UV transilluminator with minimal exposure time to prevent damage to
the DNA. Excised bands were incubated at 50°C with 3 volumes of Gel
Solubilisation Buffer (L3) for every 1 volume of agarose until the agarose was
completely dissolved. For optimal DNA yields, 1 gel volume of isopropanol was
added to the dissolved gel slice before adding to the PureLink®
Clean-up Spin Column. The DNA within the sample bound to a silica membrane of
the column by spinning at 16,000 xg for 1 minute. The sample was then washed
with an ethanol-containing buffer to remove salts and impurities. A buffer-free
centrifuge spin of 16000xg for 2 minute was included to remove residual ethanol and
buffer that may interfere with downstream steps. Finally, each sample was incubated
at room temperature for one minute with 50µL nuclease free water and eluted by
spinning at 16,000 xg for 1 minute. If not required for immediate use, both excised
bands and eluted DNA could be stored for up to one week at -20°C.

2.2.5.5  PCR clean-up procedures

All PCR products that were not extracted directly from agarose gels underwent an
additional PCR clean-up step to remove unincorporated primers and dNTPs and to
improve the fidelity of sequencing reactions. This step was carried out using
EXOsap-IT (Fisher Scientific, UK), which is a combination of Exonuclease I to
degradate single stranded DNA and Shrimp Alkaline Phosphatase to dephosphorylate
excess dNTPs. 2µL of EXOsap-IT was incubated with 5µL second round PCR
product at 37°C for 15 minutes. The active enzymes were then denatured at 80°C
for 15 minutes. The resulting mixture in theory contained only PCR product,
nucleosides and inorganic phosphate and was then used directly for sequencing
reactions.

2.2.5.6  Sanger sequencing

All sequencing undertaken during the course of this work employed the Sanger
Sequencing method, whereby the reaction mixture contains all four chain terminating
dideoxynucleotides labelled with distinct fluorescent dyes allowing the DNA sequence to be determined sequentially. This was performed using the ABI BigDye Terminator kit (Applied Biosystems, Warrington, UK), with reagents and cycling conditions outlined below (Table 2.14 and Table 2.15). PCR products were sequenced in both sense and antisense orientation using second round PCR primers. The amount of BigDye used in each reaction was adjusted according to fragment size, with up to 2µL being used for the largest sequence fragments.

Reactions products were sent to the in-house sequencing facility at The Genepool (Ashworth Laboratories, Kings Buildings, Edinburgh) and results were returned in FASTA format with corresponding chromatograms.

### 2.2.6 Analysis of Equine Liver enzymes

Biochemical analysis of liver enzymes is imperative step of both liver disease and liver failure, and can be helpful is narrowing the diagnosis, designing the most efficient treatment method and improving the ultimate prognosis. Sorbital dehydrogenase (SDH) and gamma glutamyltransferase (GGT) are liver specific enzymes that respectively can reflect hepatocellular and biliary injury. In addition aspartate aminotransferase (AST) and alkaline phosphatase (AP) can indicate hepatocellular and biliary injury, but are not strictly liver specific. Damage to hepatocytes causes AST to elevate and may also be an indicator of muscle damage. Alanine aminotransferase (ALT) often used in the diagnosis of liver disease in small
animals does not leak at significant levels from equine hepatocytes so this enzyme is not a useful diagnostic tool for hepatic disease in horses.

Abnormal liver enzymes occur when 60-70% of the liver function is lost in the horse and will likely indicate increased levels of bilirubin, ammonia, and serum iron and gamma globulins in the case of chronic disease. Therefore it is important that when studying liver disease in horses to observe a combination of liver enzymes and to apply careful interpretation of the results as discussed in Section 1.6.1. In all studies presented in this thesis GGT, GLDH and bile acids were the liver enzyme tests carried out (Chapter 4, 5)

For the purpose of all studies presented within this thesis GGT levels >42U/L, GLDH values >12U/L and bile acid levels >12µmol/l were considered as indicative of underlying hepatic insufficiencies, although only when enzyme elevations were observed in two or all enzymes measured.

2.2.7 Generation of RNA Transcript Controls

2.2.7.1 RNA Transcription

RNA transcript controls were generated for NS3 region of NPHV and EPgV from PCR positive viraemic horses. Respective regions were PCR amplified from positive samples, purified products were inserted into pGem-tEasy vector (as per manufacturers instructions Table 2.16) and PCR amplified using M13 forward and reverse primers to generate amplicons with required T7 promoter site. Sequence fidelity was verified by Sanger sequencing of the M13 PCR products (Appendix D). Reaction reagents (Table 2.17) were assembled at room temperature and incubated overnight at 37°C. Overnight for 12 hours was determined as the optimum duration to achieve maximum yield of RNA for transcripts of less than 500nt. Reaction was terminated by the addition of 1ul TURBO DNase and incubated at 37°C for 15 minutes. RNA can be subsequently purified by relevant method including Lithium
Chloride purification and Phenol: chloroform extraction followed by isopropanol precipitation.

**TABLE 2.16: Reaction components of pGem T-Easy vector system incubated overnight at 4°C**

<table>
<thead>
<tr>
<th>Reagent</th>
<th>Volume</th>
</tr>
</thead>
<tbody>
<tr>
<td>2X ligation buffer</td>
<td>5µl</td>
</tr>
<tr>
<td>pGem-T Easy (50ng)</td>
<td>1µl</td>
</tr>
<tr>
<td>PCR Product (3:1 Product/Vector Ratio)</td>
<td>Xµl*</td>
</tr>
<tr>
<td>T4 DNA Ligase</td>
<td>1µl</td>
</tr>
<tr>
<td>Nuclease Free Water</td>
<td>Up to 10µl</td>
</tr>
</tbody>
</table>

* Volume adjusted to appropriate 3:1 DNA to Vector ratio

**TABLE 2.17: Reaction components of MEGAscript T7 Kit for generation of RNA transcripts**

<table>
<thead>
<tr>
<th>Reagent</th>
<th>Volume</th>
</tr>
</thead>
<tbody>
<tr>
<td>ATP solution</td>
<td>2µl</td>
</tr>
<tr>
<td>CTP solution</td>
<td>2µl</td>
</tr>
<tr>
<td>GTP solution</td>
<td>2µl</td>
</tr>
<tr>
<td>UTP solution</td>
<td>2µl</td>
</tr>
<tr>
<td>10 X reaction buffer⁴°</td>
<td>2µl</td>
</tr>
<tr>
<td>Template DNA</td>
<td>1µg</td>
</tr>
<tr>
<td>Enzyme Mix</td>
<td>2µl</td>
</tr>
<tr>
<td>Nuclease Free Water</td>
<td>Up to 20µl</td>
</tr>
</tbody>
</table>

2.2.7.2 Lithium Chloride RNA purification

The MEGAscript kit allows for efficient purification of RNA from high yield transcription reactions. Supplied reagents allow for the removal of excess nucleotides, short oligonucleotides, salt and proteins from RNA. Lithium chloride (LiCl) precipitation removes unincorporated nucleotides and proteins, although it is not effective for the purification of RNA transcripts shorter than 300nt. Briefly RNA is precipitated by the addition of 30µl each nuclease free water and LiCl precipitation

---

⁴° 10 X Reaction Buffer contains salts, buffer, dithiothreitol, and other ingredients not detailed by supplier (Life Technologies)
solution. Following mixing, the reaction was incubated at -20°C for ≥ 30 minutes. Centrifuging at 4°C for 15 minutes at maximum speed pelleted the RNA. The supernatant was carefully aspirated and the pellet washed once with 1ml of 70% EtOH and re-centrifuged at maximum speed to remove unincorporated nucleotides. The EtOH was carefully removed and the RNA pellet resuspended in RNA storage solution\textsuperscript{11}. RNA concentration was then determined by nanodrop and the RNA stored at -80°C.

2.2.8 Recombinant viral protein production and Enzyme Linked Immunoabsorbant Assays

2.2.8.1 PCR amplification

Nested PCR primers for NPHV Core and NS3 helicase and EPgV NS3 helicase were designed based on published sequences with appropriate restriction sites for subcloning added to the inner primer sets (Appendix A). PCRs were performed on positive serum using GoTaq (Promega) in accordance with the manufacturers’ instructions, using conditions previously outlined (Table 2.3 and 2.4). The amplified fragments were then gel extracted and purified as outlined in 2.2.4.4. All purified PCR amplicons were sequenced to ensure sequence fidelity and that no mis-incorporated bases were present.

2.2.8.2 Restriction Digest

Restriction sites contained in the nested PCR primers were designed to complement the sites available in the chosen plasmids, pGex-2t (GE Life Sciences) and pET-28B (Novagen). pGex-2t produces a GST fusion protein designed to increase the solubility of the recombinant protein and reduce the production of insoluble recombinant protein in bacterial inclusion bodies. pET-28B generates C-terminal His tag fusion protein; the 6XHis tag is small and suitable for production of low molecular weight recombinant proteins. The vector and the purified PCR products

\textsuperscript{11} RNA storage solution contains RNA storage solution containing 1 mM sodium citrate at pH 6.4 ± 0.2 (Life Technologies), RNAsin at 40U/µl (Promega), Carrier RNA at 1µg/µl (Qiagen).
were digested with BamHI and EcoRI (Promega) in a double digest as per the outlined manufacturer’s instructions (Table 2.16) at the appropriate conditions (Table 2.17). The excised bands were then recovered by gel extraction and purified using PureLink™ Quick Gel Extraction and PCR Purification Combo Kit (Life technologies, Glasgow, UK) (2.2.2.4). Where restriction digests were carried out using different enzymes, the reaction conditions, enzymes and reasoning are outlined in Table 2.18.

### TABLE 2.18: Restriction Digest Reagents

<table>
<thead>
<tr>
<th>Reagent</th>
<th>Volume</th>
</tr>
</thead>
<tbody>
<tr>
<td>10X Buffer H(^{12})</td>
<td>2 µl</td>
</tr>
<tr>
<td>BSA (10 µg/µl)(^{13})</td>
<td>0.2 µl</td>
</tr>
<tr>
<td>Bam HI (10U/µl)</td>
<td>1 µl</td>
</tr>
<tr>
<td>EcoRI (10U/µl)</td>
<td>0.8 µl</td>
</tr>
<tr>
<td>DNA</td>
<td>1 µg/µl</td>
</tr>
<tr>
<td>Nuclease Free H₂O</td>
<td>To 20 µl</td>
</tr>
</tbody>
</table>

### TABLE 2.19: Restriction Digest Conditions

<table>
<thead>
<tr>
<th>Temperature</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>37°C</td>
<td>2 hours</td>
</tr>
<tr>
<td>65°C</td>
<td>15 minutes</td>
</tr>
</tbody>
</table>

### TABLE 2.20: Modifications to Restriction Digest Reagents and Conditions

<table>
<thead>
<tr>
<th>Experiment</th>
<th>Modification to reaction conditions</th>
<th>Modifications to reaction reagents</th>
<th>Relevant Thesis Chapter</th>
</tr>
</thead>
<tbody>
<tr>
<td>HBV chimpanzee sub-species typing</td>
<td>37°C 1 hour 80°C 20 minutes (Inactivation)</td>
<td>5 µl Buffer 4 (Promega) 1 µl Himf I/ Psi I</td>
<td>Chapter 3</td>
</tr>
<tr>
<td>pGex 2t Vector</td>
<td>None</td>
<td>0.75 µg/µl DNA</td>
<td>Chapter 5/6</td>
</tr>
<tr>
<td>NPHV Core Cloning</td>
<td>None</td>
<td>4 µl Buffer H 0.4 µl BSA 2.5 µl Bam HI 2 µl EcoRI Nuclease free water up to 50 µl</td>
<td>Chapter 5/6</td>
</tr>
<tr>
<td>pET28B Vector</td>
<td>None</td>
<td>0.8 µg/µl DNA</td>
<td>Chapter 5/6</td>
</tr>
<tr>
<td>EPGV NS3 Cloning</td>
<td>None</td>
<td>1.5 µg/µl DNA</td>
<td>Chapter 5/6</td>
</tr>
</tbody>
</table>

\(^{12}\) 1X Buffer H contains 90 mM Tris-HCL, 10 mM MgCl₂, 50 mM NaCl adjusted to pH 7.5

\(^{13}\) All restriction enzymes are supplied with 10 mg/ml Acetylated BSA, used at a final concentration of 0.1 mg/ml.
2.2.8.3 Ligation of Expression Plasmid and Insert

PCR amplicons of NS3 helicase of NPHV and EPgV digested and purified were cloned into the pGex-2T Vector for production of Glutathione-S Transferase fusion protein (Amersham Pharmacia Biotech). The purified core amplicon of NPHV was cloned into pET28B vector (Merck Novagen) to produce 6 X His-tag fusion proteins. T4 DNA Ligase is an enzyme that catalyses the joining of two strands of DNA in an ATP dependent manner between the 5'-phosphate and the 3'-hydroxyl groups of adjacent nucleotides in either a blunt ended or cohesive-ended configuration. Optimum ligation occurs with insert: vector ratio of between 3 and 6, and all cohesive end ligations used in this thesis were carried at 14°C overnight and transformed into JM109 high efficiency competent cells (Agilent) for propagation. JM109 cells were used as they contain endA1–DNA specific endonuclease 1 mutation- that improves quality of plasmid DNA isolation and recA1 mutation that increases stability of the inserts. Recombinant clones were selected by antibiotic resistance and were confirmed by PCR and sequencing. High concentrations of plasmid were purified from overnight cultures of ampicillin resistant colonies using Wizard® Plus SV Minipreps DNA Purification System as per manufacturer’s instructions (Promega).

**TABLE 2.21: Ligation reaction reagents used in the construction of recombinant plasmids containing NPHV and EPgV viral protein coding sequences.**

<table>
<thead>
<tr>
<th>Reagent</th>
<th>Volume</th>
</tr>
</thead>
<tbody>
<tr>
<td>10X T4 Ligase Buffer</td>
<td>2.0ul</td>
</tr>
<tr>
<td>10X ATP Solution</td>
<td>2.0ul</td>
</tr>
<tr>
<td>Vector</td>
<td>As Required</td>
</tr>
<tr>
<td>Insert</td>
<td>As Required</td>
</tr>
<tr>
<td>T4 Ligase Enzyme</td>
<td>1.5ul</td>
</tr>
<tr>
<td>Nuclease Free H₂O</td>
<td>Up to 20ul</td>
</tr>
</tbody>
</table>
Materials and Methods
FIGURE 2.3: Vector maps of pGex-2t and pET28B plasmids used in the generation of NPHV and EPgV recombinant proteins. (A) PCR amplified regions of NPHV NS3 helicase and core regions with BamHI and EcoRI sites were digested and ligated into complementary digested pGex-2t vector and used in the transformation of chemically competent BI21(DE3) E. coli to produce GST tagged NPHV recombinant proteins for application in ELISA screening. (B) The PCR amplified EPgV-NS3 helicase region was digested with BamHI and EcoRI sites and ligated into complementary digested pET28B vector and used in the transformation of chemically competent BI21(DE3) E. coli to produce C-terminal 6 X His tagged EPgV recombinant protein for ELISA and western blot.

2.2.8.4 Recombinant Protein Expression

One Shot® BL21 (DE3) pLysS Chemically Competent E. coli (Invitrogen) were transformed with plasmid as per manufacturer’s instructions. From overnight cultures of antibiotic resistant colonies a fresh 1:10 dilution was prepared and grown to OD 0.6 before induction at 20°C for 3 hours with 1mM IPTG. LB medium was removed by centrifugation at 2000 xg for 15 minutes and cell pellets were weighed and lysed with BugBuster Master Mix (Merck Millipore) containing Roche complete ULTRA protease inhibitors, as per manufacturer’s instructions. Cell lysate was recovered by centrifugation at 21000 xg for 15 minutes at 4°C. Soluble and insoluble fractions were analysed by SDS-PAGE for the presence of recombinant protein.

The optimum temperature and duration for protein expression was determined by carrying out a series of experiments measuring the level of protein expression at temperatures between 37°C and 20°C and over periods of induction between 1 and 24 hours. The optimum conditions of 20°C and 3 hours were chosen based on the observation that they produced the highest levels of soluble protein, compared to induction at higher temperatures, for example 37°C and for 3 to 24 hours which led to higher levels of detectable protein expression in the insoluble cell lysate (Section 6.2.2.3).

Purification

Glutathione Sepharose Trap 4B columns (GE Life Sciences) are designed for the recovery and purification of glutathione S-transferase (GST) tagged proteins produced using the pGEX expression vectors. Prior to proceeding to use soluble cell lysate in ELISAs and western blots, GST-Trap 4B (5ml) columns were used to try to
recover a sufficient quantity of pure NPHV-NS3 and core proteins expressed using the pGEX-2t vector. These columns allow for the direct purification of proteins from pre-treated bacterial lysates. The column is packed with glutathione sepharose and when the lysate containing the GST tagged recombinant protein is applied to the column the GST tagged proteins bind to the ligand (Glutathione and 10-carbon linker arm at a concentration of 7-10µmol glutathione/ml medium and binding capacity of approximately 25mg recombinant glutathione S-transferase /ml medium). Subsequently the GST tagged protein of interest can be released and eluted or cleavage of the bound GST using a site-specific protease (Thrombin) releases the purified protein.

Using manufacturers specifications soluble NHPHV-NS3 cell lysate (GST tagged) was centrifuged and filtered through 0.45µm filter prior to applying to the column to prevent clogging the column. The column was equilibrated with 5 volumes of binding buffer at flow rate of 1ml/min before applying the sample at flow rate of between 0.5-2ml/min (5ml column)\textsuperscript{14}. The column was then washed with 10 column volumes of binding buffer at a flow rate of 5ml/min. A solution of thrombin at 1U/µl was prepared in 500µl of cold PBS, and 400µl was added to 4.6ml PBS and loaded onto the column ensuring that the column was then sealed with the bottom and top plugs supplied. Incubation times between 2-16 hours were tested and the protein eluted using 15ml PBS collected in 1ml fractions. NPHV-NS3 protein was not detectable in the eluate by SDS page or subsequent western blot using NPHV positive serum except after a 16hour thrombin cleavage where very faint levels of reactivity were observed. Based on the low level of purified protein recovered it was considered to be an inefficient method by which to recovery the purified proteins of interest and therefore soluble cell lysates were instead applied in all subsequent

\textsuperscript{14} Binding Buffer: PBS, pH 7.4 (140 mM NaCl, 2.7 mM KCl, 10 mM Na2HPO4, 1.8 mM KH2PO4, pH 7.4)
ELISAs and western blots with appropriate pre-incubation procedures as outlined in section 2.2.8.5 and appropriate controls

2.2.8.5 Enzyme Linked Immunoabsorbant Assay (ELISA)

High-bind 96-well enzyme-linked immunosorbent assay (ELISA) plates (Greiner Bio-One) were coated overnight with BugBuster (Merck Millipore) cell lysates (0.5 µg of recombinant protein in 100 µL of carbonate buffer) or an equivalent volume of protein isolated from cells infected with non-recombinant pGex-2t collected and processed in parallel with the NS3 and Core proteins. Before carrying out large scale screening a design of experiment was conducted in which the antibody concentrations, serum concentrations and finally the percentage BSA were altered and all combinations assayed to determine the optimum ELISA conditions. Optimum conditions were considered to be the combination of parameters that produced the best ratio of signal in the test wells versus control wells. Test serum and species-specific antibody were tested at dilutions between 1:10 and 1:20,000, and 1:500 and 1:20,000 respectively, and each of these combinations was further tested using BSA concentrations between 1-5% for blocking and antibody dilution. It was determined that the best signal in the sample wells combined with the lowest level of background in the control wells was achieved using a combination of sample serum diluted at 1:1000 with a 1:4000 dilution of the HRP conjugated species-specific antibody in 3% BSA. No advantage was observed in using BSA concentrations exceeding 3% during blocking and all subsequent screening (detailed below) used a 3%BSA/PBS (wt/vol) solution.

Antigen coated high-bind 96-well plates were washed with 250 µL of 1% Tween 20/PBS (vol/vol), and coated wells were blocked with 150 µL of 3% BSA/PBS (wt/vol) at room temperature for 2 h. After 1 round of washing, test serum samples diluted to a final concentration of 1:1000 in 100 µL of 3% BSA/PBS (wt/vol) and pre-incubated with 10 X non-recombinant pGex2t cell lysate and were added to the
100

wells and incubated for 1 h. The wells were washed 6 times with 250 µL of 1% Tween 20/PBS (vol/vol), were incubated each time for 15 min, and then were incubated for 30 min with 100 µL of horseradish peroxidase–conjugated goat anti–horse immunoglobulin G antibody (Serotec) or the appropriate species specific anti-IgG diluted 1:4000 in 3% BSA/PBS (wt/vol). After 4 rounds of washing, plates were developed by adding 70 µL of the horseradish peroxidase substrate (2,2′-azino-bis 3-ethylbenzothiazoline-6-sulfonic acid; Liquid Substrate System) to each well in accordance with the manufacturer’s protocol (Sigma). Plates were allowed to develop for 10–35 min and were read at 405 nm. The immunoreactivity of serum samples to control wells was compared and the mean control plus 3 times the standard deviation was established as cut off. A second ELISA targeting NPHV Core antigen confirmed samples identified as NPHV NS3 antibody positive. EPgV antibody positive samples were confirmed by western blot analyses.

2.2.8.6 Antibody detection by Western Blot

Western blots were carried out to confirm all EPgV seropositive cases in the absence of a second confirmatory ELISA for EPgV. Whole-cell lysate and enriched protein samples were analysed on 10%/16% sodium dodecyl sulphate (SDS)–polyacrylamide gel electrophoresis (PAGE) gels. The gels were either stained with Coomassie blue (Sigma) or blotted onto Protran BA 85 nitrocellulose membranes (Whatman) by use of a semidy electroblotter. Membranes were blocked overnight in 4% Milk Powder (MP)/0.05% Tween 20/TBS (wt/vol/vol) (Marvel Milk Powder, <1% Fat). The membranes were then incubated at room temperature for 1 h 30 mins with serum obtained from horses identified as EPgV seropositive and diluted 1:2000 in 4% MP/0.05% Tween 20/TBS (wt/vol). After 6 washes in 0.05% Tween 20/TBS (vol/vol) for a total of 2 hours, the membrane was incubated with horseradish peroxidase–conjugated goat anti–horse immunoglobulin G antibody (Serotec) diluted 1:35000 in 4% MP/0.05% Tween 20/TBS (wt/vol). After 4 washes in 0.05% Tween 20/TBS (vol/vol) for a total of 1 h, bound antibody was visualized using
electrochemiluminescent prime western blotting detection reagent incubated for 5 minutes on the membrane in accordance with the manufacturer’s protocol (Amersham).

2.3 Computational Methods

2.3.1 Statistical methods

Unless otherwise stated, all statistical tests carried out on epidemiological data were two-tailed Fishers exact tests, appropriate for small datasets with limited categories for each variable.

2.3.2 Sequence alignment

All phylogenetic and evolutionary analyses were entirely dependent on the construction of a sequence alignment in which homologous sequences were correctly identified and aligned. The correct execution of the alignment process is central to all downstream analysis and prevents false conclusions regarding phylogeny, genetic diversity and recombination.

Sequences were imported into SSE v1.0 (Simmonds, 2012) and initially aligned with a combination of the CLUSTAL-W or MUSCLE version 3.8 algorithms (Thompson et al., 1994, Edgar, 2004) implemented within SSE v1.0. Alignments were refined where absolutely necessary by manual editing at the amino acid level (for coding sequences) and nucleotide level (for non-coding sequences). PCR products were sequenced in both sense and antisense directions to allow for dual coverage. Any gaps and mismatches between the sense and antisense sequences were resolved by inspection of accompanying chromatograms in the Finch TV Version 1.4.0 software (www.geospiza.com./finchtv). Amino acid sequences were obtained by translation of nucleotide sequences carried out in the SSE v1.0 software package (Simmonds, 2012) and using a standard genetic code.
2.3.3 Database searching

A critically important step in many of the analyses within this thesis was the construction of comprehensive datasets containing all published sequences of a particular gene region or species. The inclusion of all available published sequences with sequences generated from our studies enable more comprehensive investigations to be carried out. All database searches were carried out on nucleotide sequences available in the three major online databases (Genbank NCBI, European Molecular Biology Laboratory (EMBL) and DNA Databank of Japan (DDBJ)). The agreement between all three major databases to cross-submit all sequences allowed all necessary information to be accessed via the NCBI Basic Local Alignment Search Tool (BLAST).

The BLAST method of searching sequence databases was first described in 1990 (Altschul et al., 1990) and is based on locating regions of local similarity between a submitted query sequence/sequences and all database entries. This procedure was then extended in 1997 to an algorithm more based in biological reality due to its ability to deal with gaps in sequence alignments (Altschul et al., 1997). BLAST uses a heuristic method to identify similar sequences, this focuses on locating short regions of sequence similarity, which while fast and reasonably accurate does sacrifice some of its accuracy for speed and cannot guarantee that all homologous sequences will be returned. The optimization of certain parameters is therefore critical to the accuracy of the endeavour.

All studies presented within this thesis utilised a nucleotide BLAST procedure to obtain the maximum number of matching and potentially homologous sequences for analysis. In brief, an input query sequence in FASTA format is broken down into nucleotide fragments of a certain length. The entire database is then searched for occurrences of every possible fragment generated from the query sequence. These have a default length of 11 nucleotides. However, as an exact match of the entire
Materials and Methods

fragment length is necessary for progression of the algorithm, the length of fragment can be specified in order to adjust the sensitivity and specificity of the protocol. When a match is located in a database sequence, adding bases from the query sequence at both the 5’ and 3’ ends extends the hit and checks for matches in the database sequence. The growing alignment is sequentially scored based on a specified match/mismatch penalty. The default value for match/mismatch score is 1/-2. However, this was decreased to 1/-1 for searches for sequences that were more distantly related. The alignment extension is continually scored until the score drops below a specified threshold (20 for nucleotide sequences). The extension step allows the differentiation of meaningful matches from random ones. Sequences returned were assigned an E-value, which is an estimate of the probability of a hit being a false positive.

2.3.4 Construction of phylogenetic trees

Phylogenetic analysis uses sequence information to provide an estimate of the evolutionary relationships between sequences, provided that adequate homology exists. Studies within this thesis were largely reliant on the use of the Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0 (Tamura et al., 2007) and latterly, version 5.2 (Tamura et al., 2011). Unless otherwise stated, all phylogenetic trees used for analysis of HBV, HCV and HPgV homologs were constructed by neighbour-joining method (Saitou and Nei, 1987) from 100 bootstrap re-sampled sequence alignments of maximum composite likelihood (MCL) (Guindon and Gascuel, 2003) distances with pairwise deletion for missing data. Phylogenetic trees were improved by the inclusion of an outlier, which rooted the tree and therefore provided an estimate of which nodes under study were the most ancient (sequences used to root phylogenetic trees are specified throughout).

---

15 Phylogenetic trees included in Chapter 5 were constructed with 500 bootstrap repetitions.
However, for analyses of less genetically diverse sequences, a sequence of different species or type was used (specified throughout).

The neighbour-joining (NJ) method was first described by Saitou and Nei in 1987 (Saitou and Nei, 1987). It is a distance based method that produces an additive tree and is one of the most widely used methods for phylogenetic tree reconstruction (Gascuel and Steel, 2006). While the authors have stated that the method is not guaranteed to always produce the minimum evolution tree, it has good reliability in producing the correct tree topology (Saitou and Nei, 1987). The fact that it is significantly more computationally efficient than more traditional methods based on maximum parsimony allows the relatively fast calculation of trees reconstructed from a large number of operational taxonomic units (OTUs). As the name suggests, NJ is based on the principle of sequentially joining the most closely related OTUs via a single interior node in an unrooted bifurcating tree. Initially, all OTUs/sequences are represented in a star-like tree with no internal clustering. Evolutionary distances are calculated in a pairwise fashion between all sequences and the OTU pair with the shortest inferred branch lengths are joined by an internal tree node. The algorithm then considers this pair as a single OTU and repeats the process of analysing evolutionary distances between all sequences, culminating in the joining of the pair to their nearest neighbour. This process is repeated sequentially until all sequences are included in the final tree.

One of the limitations of the NJ method is that it assumes that no backwards or parallel substitutions have taken place. Therefore, NJ trees are subject to long branch attraction (Felsenstein, 1978), whereby it is possible to predict that two OTUs are more closely related than they really are, as the algorithm cannot correct for the occurrence of convergent substitutions at specific sites. Other phylogenetic reconstruction methods such as those relying on maximum likelihood optimality criterion are generally more biologically accurate due to their greater ability to
correct for multiple hits (Holder and Lewis, 2003). However, as maximum likelihood methods of tree-building seek to explore a significant proportion of the tree space and score each tree, they are very computationally demanding methods and not practical for use on large datasets.

The dual problems of inaccurate phylogeny estimation and long branch attraction exhibited in pure NJ algorithms can be prevented by the usage of an MCL method to calculate evolutionary distances within a dataset (Tamura et al., 2004). The calculation of MCL distances is ideal for rapidly evolving sequences, as it automatically corrects for multiple mutational events at the same site and additionally incorporates a model of nucleotide substitution that is optimized automatically to fit the data (Tamura et al., 2011). This includes such aspects as transition/transversion ratios, where it is acknowledged that purine-purine and pyrimidine-pyrimidine substitutions occur more readily than purine-pyrimidine in natural mutations.

2.3.5 Bootstrapping

In order to determine the robustness of observed phylogenetic groupings within the tree, all trees were constructed with either 100 or 1000 bootstrap resamplings. The bootstrap is a statistical technique that was first applied to assessing confidence levels in phylogenetic groupings by Professor J. Felsenstein of the University of Washington in 1985 (Felsenstein, 1985). The generation of bootstrap values for internal nodes of a phylogenetic tree gives an estimate of the reproducibility of a monophyletic clade. Sequence data used to construct the original tree is randomly resampled and trees constructed from each of the pseudo-replicates produced. The proportion of observed trees in which each monophyletic clade is observed is then computed and reported as the bootstrap value (Holder and Lewis, 2003). A high bootstrap value (close to 100) indicates that analysis of randomly selected subsets of the alignment data consistently predict the same clade (Berry and Gascuel, 1996).
One of the major limitations of bootstrap analysis is that it fails to provide an estimate of the probability of an observed tree and merely indicates a degree of robustness for the tree phylogeny. For instance, due to the inherent bias in methods of calculation of evolutionary distances, long branches subject to long branch attraction can have good bootstrap support, even if they do not represent a “true” phylogenetic relationship (Swofford et al., 2001).

For all phylogenetic studies included in this thesis, a bootstrap value of $>70\%$ was used as the threshold for determining adequate support. MEGA v4.0 and v5.0 software packages present phylogenetic trees as both a majority rule bootstrap consensus tree and as the originally determined tree with bootstrap values superimposed. The bootstrap consensus tree was used for phylogenetic analyses.

### 2.3.6 Recombination detection methods

The general strategy for the detection of recombination in sequence alignments was multi-faceted. Central to all analysis undertaken was the topology of bootstrap supported phylogenetic trees (Chapter 3). Phylogenetic trees were constructed containing sequence fragments of varying sizes (depending on the specific analysis undertaken). Trees from sequential genome regions were manually inspected to identify phylogeny violations between clades with bootstrap support above the threshold of $>70$. All recombination events detected from the inspection of phylogenetic trees were further characterised by analysis with GroupScan and where possible, putative breakpoints were determined.

In addition to the initial screening of sequence datasets by phylogenetic reconstruction, several computational analyses were undertaken to effectively screen each dataset for the occurrence of recombination. This was done to supplement and support results obtained from phylogenetic analysis. Screening analyses were undertaken SSE v1.0 (Simmonds, 2012) All algorithms used for the detection and description of putative recombination events are discussed below.
2.3.6.1 **TreeOrder Scan**

The program TreeOrder Scan in SSE v1.0 is a sequence analysis tool used to directly visualize the frequency of recombination events in a particular sequence alignment and provide a preliminary analysis of potential breakpoints. TreeOrder Scan analyses a sequence alignment over sequential fragments of defined length and increment. Optimal trees are produced for each sequence fragment and branches collapsed to a pre-specified bootstrap support threshold. The program then directly compares the branching order of trees from each fragment and catalogues any phylogeny violations. This is then presented as a half-diagonal matrix, giving an overview of the degree of phylogenetic incongruence observed in each region and additionally as a specific list of putative violations. Sequence datasets can be sub-divided into groups, which then allows the computation of frequencies of phylogeny violations both within and between groups (for instance, different HBV NHP species).

TreeOrder Scan was used as an initial tool to screen alignments for recombination, using a sequence fragment of 300 nucleotides and an increment of 30. This was implemented in analysis of recombination within the HBV genome of gorilla isolate (Chapter 3).

2.3.6.2 **GroupScan**

In instances where recombination was detected by inspection of phylogenetic trees and the usage of the computational methods described above, GroupScan was used in order to verify putative events and determine recombination breakpoints where possible. GroupScan is also implemented within the SSE v1.0 software package.

GroupScan was first described in 2005 (Simmonds *et al.*, 2005) and is used to score the extent of grouping of a query sequence with two or more control groups. The grouping score is a measure of how deeply embedded within a particular group a query sequence is. For instance, a score of 0 indicates that there is no grouping and a
score of 1 indicates that the query sequence groups deep within a clade of control sequences. Sequences are analysed in a sequential fashion, with both fragment size and increment being specified by the user. The output shows a graph with the midpoint of the analysed fragment represented on the x-axis and the grouping score on the y. This can be used to determine approximate recombination breakpoints, when a query sequence is scanned against its nearest neighbours in both regions. The putative breakpoint is determined by point of intersection of the grouping score trace of both parental sequences.

The GroupScan method was used for all determinations of recombination breakpoints. In preference to the method commonly employed in BootScan and Simplot analyses, GroupScan can consider entire groups of sequences as control groups and therefore doesn’t lose valuable phylogenetic information by condensing these sequences to a single consensus. This also allows the calculation of recombination breakpoints in circumstances where parental sequences are not present within the dataset analysed.

Within this thesis, GroupScan was used to pinpoint recombination breakpoints in regions of the gorilla HBV genome and to confirm previously identified recombinants between human HBV genotypes and humans and chimpanzee HBV species (Chapter 3). All analyses carried out used a fragment size of 300 nucleotides and an increment of 30, unless otherwise stated.

2.3.7 RNA secondary Structure Analysis

2.3.7.1 Structure Dist

As part of the sequence analysis software SSE v1.1, StructureDist analyses a group of aligned sequences and generates the most energetically favoured RNA structure. UNAFOLD computes minimum energy structures for each sequence in the alignment file and sequence pairings are compared to identify conserved and non-conserved structural predictions between sequences. This simple method of structural prediction
Materials and Methods

takes into account the naturally occurring sequence variability of the aligned sequences, while assuming the presence of underlying conserved RNA folding, to determine and refine the secondary structure prediction of the RNA. StructureDist is complimentary to the PFold stochastic context-free grammar and covariance detection weighting method that also acknowledges the presence of sequence variability (Knudsen and Hein, 2003, Gruber et al., 2007).

For each site of the sequence alignment, a pairwise comparison between two structure predictions can have the following outcomes:

- **PPS (Paired, Paired, [Same])**: Predicts that for two files there is a site x, with a predicted pairing to a downstream site, y, and that both structural files have identified the same prediction.

- **UUS (Unpaired, Unpaired, [Same])**: Predicts that for both sequences, site x is unpaired.

- **PPD (Paired, Paired, [Different])**: Predicts site x as base-paired in both structure files, but the downstream bases, y1 and y2 are different.

- **UPD (Paired, Unpaired, [Different])**: Predicts that in the two structural files one of the bases at site x is predicted to be base-paired, while the corresponding base in the second file in unpaired.

From the StructureDist settings menu the program can be set to produce four different output formats.

1) **Pairing scan output** provides a listed summary of the outcome of comparing sites for each pairwise comparison of sequences. The frequency of occurrence of PPS/UUS and PPD/UUD provides an indication of the overall degree of similarity between the structure prediction files.

2) **Position file output** lists the sums of outcomes of each pair-wise comparison of sequences at each nucleotide position in the alignment.

3) **Graph output** can be directly generated using this analysis option. Output as a matrix file provides list each site comparison (*i.e.* PPS, UUS, PPD, UPD), its
frequency, X and Y coordinates, and spacing between upstream and downstream bases. This format can be used to generate graphical representation of the results in a 2D dotplot which records the general conservation of predicted base pairings and specific pairings which are presented as colour coded dots on X, Y plot to indicate the likelihood of each pairing.

4) Loop statistics output presents a series of files .DT1 – .DT4 listing the structure elements and their degree of sequence conservation, the frequencies of each base-pair in duplex regions and of each base in unpaired regions with differing degree of structural conservation, further breakdown the frequencies listed by their relative upstream and downstream positions and the numbers, frequencies and ratio of observed pairings to those expected from base composition of the sequences.

2.3.7.2 The Vienna Server: ALIFOLD secondary structure prediction

Using the standard energy model (Hofacker et al., 2002) with the addition of a term for covariance the Alifold server predicts a consensus RNA structure from a set of aligned sequences (Gruber et al., 2008b). Sequences can be input as aligned FASTA files or Clustal format, up to 3000 nt per sequence and a maximum of 300 sequences. RNAalifold version applied in the analysis can be set by the user under the menu and includes new RNAalifold with RIBOSUM scoring, new RNAalifold and original RNAalifold. In addition the menu offers a choice of folding algorithm, MFE (minimum free energy) alone or incorporating a partition function. Parameters favouring a) the presence or absence of GU pairs at the base of helices, b) the most explanatory structure over the consensus, and c) avoiding isolated base pairs in the predicted structure, can also be pre-set by the researcher.

Results can be visualised based on operators chosen output including interactive plots of secondary structure, where partition function is applied RNA secondary structures with reliability annotation is generated, and finally mountain plots.
The alifold RNA structure prediction method is controlled by two options:

- The weight of covariance term determines the comparative strengths of the RNA secondary structure energy versus covariance term and a value of 0 to 10 is assigned.
- Non-compatible sequence penalties determine to what extent pairs that cannot be formed by all sequences are penalised within the covariance term of the algorithm.

The results page generates the consensus sequence with an optimal consensus structure in bracket notation with its energy value. A predicted postscript drawing of the optimal structure is also presented, where circles denote sequence variations supporting the structure, while two shades of grey represent pairs with non-compatible sequences. As a consequence of selecting the of the partition function additional links will be available:

- A colour coded postscript mountain plot.
- Pair probabilities represented in a postscript dot plot.
- A text file with data on each probable pair, ranked by the reliability of each pairing.

2.3.7.3 PFold

The PFold algorithm was designed as an improvement upon the work of Knudsen and Hein (Knudsen and Hein, 1999) which combining an evolutionary model of RNA sequences with a probabilistic model of structure, assumes an alignment and produces a communal secondary structural prediction for all sequences. PFold improves up on this algorithm by making it more efficient and more resilient towards errors present in sequence alignments including how the algorithm treats sequence gaps and unknown nucleotides. Sequences are input in FASTA format, with current
server limitations of 40 sequences and a maximum alignment length of 500 nt. Known structures can also be incorporated by identifying a sequence as ‘structure’ and inputting the relevant code as outlined on the server http://daimi.au.dk/~compbio/pfold/. PFold treats sequence gaps as unknown nucleotides and therefore assigns probability of 1 to each nucleotide for that gapped position. Where only partial information is available or nucleotides are unknown PFold assigns a probability of one to each of the possible nucleotides for the unknown positions. To make the program more robust PFold assigns a small probability to each nucleotide of being another nucleotide, thus allowing for an improved treatment of sequencing errors, alignment errors and structural differences.

Results are presented in three output formats:

- **A text alignment:** A series of parentheses and dots are used to display the generic common structure of all sequences, which is used to infer the individual sequence structures. Symbols are used to score the predicted pairings.

- **A dot plot:** Comparable to SSE graphical output, the dots inside the square represents pairing probabilities between sequence positions, while dots outside the graph square represents probabilities of sequences not pairing.

- **Phylogenetic tree:** Calculated from the pairwise distances between sequences using the neighbour joining algorithm and branch lengths adjusted to maximum likelihood estimates.
Chapter 3
Species Association and HBV recombination between Non-human primate variants

3.1 Introduction

3.1.1 Detection of non-human primate variants of HBV

High frequencies of active and resolved infections with species-associated HBV variants distinct from human variants have been detected in chimpanzees (Hu et al., 2001, Hu et al., 2000, MacDonald et al., 2000, Starkman et al., 2003, Takahashi et al., 2000, Makuwa et al., 2007, Vartanian et al., 2002), gorillas (Grethe et al., 2000b) and Southeast Asian gibbons and orang-utans (Warren et al., 1998, Grethe et al., 2000b, Warren et al., 1999, Verschoor et al., 2001, Sall et al., 2005, Noppornpanth et al., 2003, Sa-Nguanmoo et al., 2009) in addition to a single isolate from a woolly monkey (Lanford et al., 1998) as discussed in section 1.2.6. The 10–12 NHP taxa-associated variants are distinct from the human variants of HBV and occur at infection frequencies comparable to human rates in endemic regions (Makuwa et al., 2005, Sa-Nguanmoo et al., 2009).

Previously published studies have examined the distribution of HBV variants among chimpanzees and sought to determine a correlation between geographical or species clustering and the distribution of chimpanzee sub-species in Africa (Starkman et al., 2003, Makuwa et al., 2007). The finding of species-specific variants of HBV in Old World great apes suggests a species-specific evolution within primates over the past 20-35 million years (MacDonald et al., 2000). There is preliminary evidence to suggest that viral sequences from geographically distinct sub-species of chimpanzee form distinct phylogenetic clusters supporting the hypothesis that HBV co-speciated within chimpanzees (Hu et al., 2001). In contrast subsequent studies argue in favour of a geographical association which would account for the clustering of orangutan
variants deep within the gibbon clade, where they occupy overlapping and adjacent habitats in contrast to variants infecting *Hylobates lar* and mainland Asia gibbons which form separate groups (Starkman et al., 2003, Norder et al., 1996, Sall et al., 2005). The suggestion that the introduction of HBV into the orangutan species is a relatively recent event is supported by the observed lack of sequence diversity between isolates and also the narrow geographical distribution of the virus within the species (Starkman et al., 2003, Sall et al., 2005, Warren et al., 1999, Warren et al., 1998, Sa-Nguanmoo et al., 2009).

### 3.1.2 Evidence for recombination between genotypes of human HBV

Recombination had been previously identified in HBV, a 196bp region of the preCore/Core region was observed to enhance recombination in vitro in the presence of extracts from actively dividing cells (Hino et al., 1991). Studies looking at integrated and episomal HBV in a single patient also detected indications for homologous recombination (Georgi-Geisberger et al., 1992). In addition to the currently assigned HBV genotypes, recombination between human genotypes has been documented, for example recombination between genotypes A and D (Bollyky et al., 1996, Bowyer and Sim, 2000, Morozov et al., 2000, Owiredu et al., 2001b) and B and C (Bollyky et al., 1996, Morozov et al., 2000, Sugauchi et al., 2002b, Mukaide et al., 1992) can generate novel variants, contributing to the genetic diversity of the virus as discussed in section 1.2.7. Tentative evidence for the occurrence of recombination has been obtained between the human genotype C and the chimpanzee variant AF498266 (Magiorkinis et al., 2005) and gibbon variants (Sa-Nguanmoo et al., 2009). Their occurrence demonstrates that despite their genetic divergence, human and non-human associated variants of HBV can share hosts in nature. Certain questions still remain surrounding HBV and the role HBV recombination may play, for example, in the geographical differences observed in
Species association and HBV recombination in NHPs

pathogenicity and viral evolution and the variations in immune responses observed in infected patients (Foster and Thomas, 1993)

A recently published study characterising HBV variants infecting ape populations in Cameroon (Njouom et al., 2010) demonstrated the existence of a gorilla-specific HBV strain and evidence of recombination between HBV strains circulating in chimpanzees. This and previous studies of HBV nucleotide sequence similarity indicate non-human primates (NHP) have distinct species-specific variants of HBV distinguishable both from each other and from human HBV despite occupying overlapping geographical areas (MacDonald et al., 2000, Robertson and Margolis, 2002, Norder et al., 1996).

3.1.3 Hypothesis on the evolution and origins of HBV

It has been difficult to propose potential evolutionary hypotheses which may account for the interspersed sequence relationships between the human HBV genotypes and those infecting gorillas, and chimpanzees in Africa and gibbons and orangutans in South Asia. Over short time frames rates of sequence change between $1.5-5 \times 10^{-5}$ substitutions per site per year have been recorded (Okamoto et al., 1987, Hannoun et al., 2000a), although due to the compact genomic organisation of HBV with several overlapping reading frames, limitations are placed on our ability to model constraints on sequence change that might effect the long term diversification of the HBV.

Prior to the detection of NHP HBV, an initial hypothesis proposed HBV originated in the Americas and as a consequence of colonisation spread to the Old World (Bollyky et al., 1997). Such a hypothesis would subsequently have meant that humans were the source of NHP infections over a relatively short period of time (Simmonds, 2001b). Following the detection of infections in wild primates it has been necessary to revise the hypothesis. Studies have examined the potential that
HBV evolution corresponds to the dissemination of humans out of Africa between 100,000 and 200,000 years ago (Norder et al., 1994, Magnus and Norder, 1995). Such a proposal however would require that the genetic relationships between human populations be reflected in the phylogeny of HBV genotypes, as is the case for HPgV (Pavesi, 2001) and this is not observed for HBV. Genotype B and C infecting Mongoloids in South Asia for example, is inconsistent with the presence of genotype F in native Americans, their nearest relatives (Alestig et al., 2001, Arauz-Ruiz et al., 2002). Furthermore the theory would require that HBV in NHPs display far greater genetic diversity between species and between human and NHP HBV given the duration of co-speciation of primate species (≈15 million years).

Finally an alternative hypothesis of coevolution argues that HBV variants in gorillas, chimpanzees, gibbons, orangutans and the Woolly monkey co-speciated between 10-35 millions years. Such a hypothesis is supported by the finding of species-specific variants of HBV in diverse species such as squirrels and woodchucks that may be the result of comparable coevolution events over extended periods of time. Again there are certain limitations to this hypothesis as it does not explain the presence of the equally distinct human genotypes, or is it possible that akin to the spread of HIV-1 and HIV-2 to humans from chimpanzees and sooty mangebys respectively (Gao et al., 1999, Feng et al., 1992), that HBV genotypes are also the result of several cross-species transmission events. However to date no wild primates have been found infected with human variants. The recent detection of three novel species of hepadnavirus infecting bats in Central America and Africa has given rise to the suggestion that bats may be the natural reservoirs for hepadnaviruses and were subsequently transmitted to primates/humans through the hunting and consumption of bush meat (Drexler et al., 2013b).

Cameroon is within a region of endemic human HBV infection with a hepatitis B surface antigen (HBsAg) prevalence in humans of 8% or greater (Mbanya et al.,
Additionally, four different great ape taxa also occur in Cameroon, providing the conditions for potential inter-species transmission. Therefore the potential for cross-species transmission to occur is much higher in these regions and the probability of detecting such an event is greater. Samples from NHPs obtained for this study all originated in different areas of Cameroon and aimed to establish whether these primates were host to human HBV variants. Although no human-derived genotypes of HBV were detected in non-human primates screened in the study presented here, evidence for transmission of HBV between chimpanzee subspecies and between chimpanzees and gorillas was obtained. Host mitochondrial sequencing and phylogenetic analysis of HBV variants from chimpanzees was applied to find supporting evidence for co-speciation or allopatric speciation of HBV among non-human primates in Cameroon.

### 3.2 Materials and Methods

#### 3.2.1 Sample selection

A total of 164 non-human primate plasma samples were screened for the presence of HBV DNA. Samples were collected from animals at three wildlife sanctuaries in Cameroon between August 2004 and August 2009. Animals were brought to the sanctuaries following confiscation by the authorities or abandonment by owners. All chimpanzees and gorillas sampled were wild-born, while other species included both wild and captive-born individuals. *Gorilla gorilla* and *Pan troglodytes* were generally housed separately in the sanctuaries, while some species of monkeys were housed in mixed groups. However, the captive history of some animals is unknown as some were held in captivity prior to their arrival in the sanctuaries and the sanctuaries themselves were not always under the current management.

Blood samples were collected via venepuncture from 73 apes comprising 11 gorillas (*Gorilla gorilla*) and 62 chimpanzees (*Pan troglodytes troglodytes* and *Pan troglodytes*)...
Species association and HBV recombination in NHPs

troglodytes elliott), and from a variety of Old World Monkey species: Cercocebus agilis (n = 7), C. torquatus (n = 2), Cercopithecus cephus (n = 3), C. erythrotis (n = 4), C. l’hoesti preussi (n = 4), C. mona (n = 9), C. nictitans (n = 3), C. pogonias (n = 1), C. tantalus (n = 3), Erythrocebus patas (n = 3), Lophocebus albigena (n = 5), Mandrillus leucophaeus (n = 20), M. sphinx (n = 9), and Papio anubis (n = 20). Plasma was separated by centrifugation and frozen at −80°C until testing. Samples were shipped to the United Kingdom from Cameroon in compliance with UK and Cameroon laws and the Convention on International Trade in Endangered Species of Wild Fauna and Flora (CITES) (Sharp et al., 2010).

3.2.2 HBV screening and whole genome sequencing

DNA extractions were performed with 50 μl of plasma using the AllPrep DNA/RNA minikit (Qiagen) according to the manufacturer's instructions with DNA eluted in a final volume of 50 μl. Samples were screened by nested PCR using first round primers S1, S5 and second round primers S3 and S6 as previously published (MacDonald et al., 2000). Three μl of extracted DNA was then amplified in a PCR mixture containing Promega Access reagents (Promega, Chilworth, Southampton, United Kingdom). First-round amplification involved 30 cycles of 94°C for 18 s, 55°C for 21 s, and 72°C for 1.5 min; and 1 cycle of 72°C for 5 min. One μl of first round PCR product was amplified further in a second-round PCR using the internal primers and conditions previously described (MacDonald et al., 2000). PCR positive samples with screening primers included samples from 9 apes (n = 2 gorillas, n = 7 chimpanzees) and 2 Old World Monkeys. The entire HBV genome was successfully sequenced from 2 gorillas (ECO50003LIP3 and ECO50065) and 6 chimpanzees (ECO50083, ECO50210, ECO51109, ECO51212, ECO51394 and ECO51377) in overlapping fragments (Figure 3.1) using primers as published (MacDonald et al., 2000b) in addition to new primer sets designed as part of this study (Table 3.2)(Figure 3.1).
**TABLE 3.2: HBV primers designed in this study to generate complete genomes.** Primers were designed to generate overlapping amplicons and used in addition to previously published sets (MacDonald et al., 2000b) to generate complete HBV genomes from all HBV screen positive NHPs.

<table>
<thead>
<tr>
<th>Primer</th>
<th>Genome</th>
<th>PCR Position</th>
<th>Sequence 5’→3’</th>
</tr>
</thead>
<tbody>
<tr>
<td>49</td>
<td>375</td>
<td>Outer Sense</td>
<td>CTG GAT GTG TCT GCG GCG TT</td>
</tr>
<tr>
<td>48</td>
<td>395</td>
<td>Inner Sense</td>
<td>CCA ATT TGT CCT GGY TAT CG</td>
</tr>
<tr>
<td>50</td>
<td>1552</td>
<td>Inner Anti Sense</td>
<td>TAA AGA GAG GTG CGK CCC GT</td>
</tr>
<tr>
<td>51</td>
<td>1542</td>
<td>Outer Anti Sense</td>
<td>GCA CAG ACG GGG AGA CCG CG</td>
</tr>
<tr>
<td>52</td>
<td>1082</td>
<td>Outer Sense</td>
<td>CWT TRT ATG ATG TAT ACA AGC</td>
</tr>
<tr>
<td>53</td>
<td>1121</td>
<td>Inner Sense</td>
<td>TCG CCA AYT TAY AAG GCC TT</td>
</tr>
<tr>
<td>54</td>
<td>2033</td>
<td>Outer Anti Sense</td>
<td>GCG GTG TCR AGR AGA RCA CG</td>
</tr>
<tr>
<td>55</td>
<td>2054</td>
<td>Outer Anti Sense</td>
<td>GCC TTC MCG GTA CAR AGC TGA</td>
</tr>
<tr>
<td>56</td>
<td>2025</td>
<td>Outer Sense</td>
<td>TTG CCT KCT GAY TTC TTT CC</td>
</tr>
<tr>
<td>57</td>
<td>2052</td>
<td>Inner Sense</td>
<td>CGT GAT CTY CTY GAC ACC GC</td>
</tr>
<tr>
<td>58</td>
<td>2868</td>
<td>Inner Anti Sense</td>
<td>CAA GAA TAT GGT GAC CCA CA</td>
</tr>
<tr>
<td>59</td>
<td>2888</td>
<td>Outer Anti Sense</td>
<td>CCC ATG CTG TAG CTC TTG TTC</td>
</tr>
</tbody>
</table>

First round PCR involved 35 cycles of 94°C for 18 s, 55°C for 21 s, and 72°C for 1.5 min; and 1 cycle of 72°C for 5 min, followed by second round PCR of 40 cycles at matching conditions. Touch-down PCR between 65°C and 50°C with 0.5°C decline per cycle was also applied (Chapter Two: Table 2.13).
3.2.3 Determination of chimpanzee sub-species

Mitochondrial sequencing to confirm host species and sub-species for the 9 complete genomes was carried out using primate specific primers with degenerate bases R: A/G, Y: T/C, D: G/A/T and N: G/A/T/C (Table 3.3). First and second round PCR conditions involved 30 cycles of 94°C for 22 s, 50°C for 24 s, 72°C for 1.5 min; and 1 cycle of 72°C for 5 min. Based on the size of the amplicons and species specific nucleotide changes in sequences, the exact species and sub-species of chimpanzee was determined. Comparison of mitochondrial sequences and HBV variants was used to determine the presence or absence of species and sub-species variants of HBV among NHPs.
TABLE 3.3: Mitochondrial specific primers to determine NHP species and sub-species of each HBV positive sample. Based on the size of the PCR amplicons and the sequences generated, chimpanzee sub-species were assigned to each HBV positive NHP and compared to the HBV sub-species variant detected.

<table>
<thead>
<tr>
<th>Primer</th>
<th>PCR Position</th>
<th>Sequence 5’→3’</th>
</tr>
</thead>
<tbody>
<tr>
<td>PrCOI</td>
<td>Outer Sense</td>
<td>CTA TTY GGY GCA TGA GCN GG</td>
</tr>
<tr>
<td>PrCOI</td>
<td>Inner Sense</td>
<td>CAG CCC TAA GYC TYC TYA TTC G</td>
</tr>
<tr>
<td>PrCOI</td>
<td>Inner Anti Sense</td>
<td>GAY DGA TCA GAC RAA YAR GGG</td>
</tr>
<tr>
<td>PrCOI</td>
<td>Outer Anti Sense</td>
<td>TAR AAG AAR GTR GTR TTR AGG TTR C</td>
</tr>
</tbody>
</table>

3.2.4 Sequencing of PCR products

Positive second round PCR amplicons were sequenced in both directions using the inner sense and inner antisense primers used in the second round of amplification. Sequencing was executed using Big Dye Terminator v3.1 (Applied Biosystems) as described in Chapter 2. Sequences were read at the Gene Pool facility (University of Edinburgh) and analysed using SSE v1.0 software (Simmonds, 2012). All sequences were aligned to reference genomes, and any discrepancies between forward and reverse sequences were resolved by analysis of chromatograms using Finch TV v1.4 (Geospiza). Sequences obtained in this study have been assigned the GenBank accession numbers JQ664502–JQ664509.

3.2.5 Phylogenetic Analysis

Phylogenetic trees were constructed using a bootstrap neighbour-joining method with 100 replications and incorporating the Kimura-2-Parameter model of nucleotide substitution and a uniform rate variation among sites using the MEGA 5.01 (Tamura et al., 2011) software package with pairwise deletion for missing data. Tree construction involved two datasets of 31 complete HBV genome sequences from Pan troglodytes troglodytes (P.t.troglodytes), Pan troglodytes elliotti (P.t.elliotti), Pan troglodytes verus (P.t.verus), Pan troglodytes schweinfurthii (P.t.schweinfurthii) and Gorilla gorilla (G.g.gorilla). GenBank sequences and final phylogenetic analysis
Species association and HBV recombination in NHPs

included representative human HBV sequences of genotypes A–H. All trees were rooted using single available Woolly monkey sequence NC001896.

Table 3.4: Genbank Accession Numbers of human HBV genotypes analysed in phylogenetic analysis with NHP HBV isolates. Representative divergent HBV sequences from all currently assigned human genotypes were included in the phylogenetic analysis of the complete genome of HBV sequences from NHPs.

<table>
<thead>
<tr>
<th>HBV Variant</th>
<th>Genbank Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>AB194951</td>
</tr>
<tr>
<td></td>
<td>AB212625</td>
</tr>
<tr>
<td></td>
<td>HBV18857</td>
</tr>
<tr>
<td></td>
<td>AY373428</td>
</tr>
<tr>
<td></td>
<td>AB033554</td>
</tr>
<tr>
<td></td>
<td>HHBV18855</td>
</tr>
<tr>
<td></td>
<td>AY236162</td>
</tr>
<tr>
<td></td>
<td>AB076679</td>
</tr>
<tr>
<td></td>
<td>AB241116</td>
</tr>
<tr>
<td></td>
<td>AY220702</td>
</tr>
<tr>
<td></td>
<td>HBVP6PCXX</td>
</tr>
<tr>
<td></td>
<td>AY233277</td>
</tr>
<tr>
<td></td>
<td>AY217368</td>
</tr>
<tr>
<td></td>
<td>AY206379</td>
</tr>
<tr>
<td></td>
<td>AB090269</td>
</tr>
<tr>
<td>E00192</td>
<td>AY800392</td>
</tr>
<tr>
<td></td>
<td>AB033557</td>
</tr>
<tr>
<td></td>
<td>AF043594</td>
</tr>
<tr>
<td>E</td>
<td>X75675</td>
</tr>
<tr>
<td></td>
<td>AY090456</td>
</tr>
<tr>
<td></td>
<td>AB056513</td>
</tr>
<tr>
<td></td>
<td>AY090457</td>
</tr>
<tr>
<td></td>
<td>AB194948</td>
</tr>
<tr>
<td></td>
<td>AY311369</td>
</tr>
<tr>
<td></td>
<td>AB056515</td>
</tr>
<tr>
<td></td>
<td>AY090454</td>
</tr>
<tr>
<td></td>
<td>AB194947</td>
</tr>
<tr>
<td></td>
<td>HHVBF</td>
</tr>
<tr>
<td></td>
<td>AF160501</td>
</tr>
<tr>
<td></td>
<td>AY090460</td>
</tr>
<tr>
<td></td>
<td>AP007262</td>
</tr>
<tr>
<td></td>
<td>ABI16550</td>
</tr>
<tr>
<td></td>
<td>ABO59659</td>
</tr>
<tr>
<td>X75664</td>
<td>ABI16850</td>
</tr>
<tr>
<td></td>
<td>ABO59661</td>
</tr>
</tbody>
</table>

3.2.6 Recombination Analysis

3.2.6.1 Tree Order Scan

Recombination analysis was carried out using Tree Order Scan package of SSE v1.0 (Simmonds, 2012) (Chapter 2.3.6.1) generating an image of individual sequence positions in phylogenetic trees generated from sequential 250-base sequence fragments, incrementing by 25 bases. Changes to the sequence order due to changes in phylogeny at the 70% bootstrap level are reported. The position of changes in the tree order of an alignment indicates the forced re-grouping of sequences into different, bootstrap supported clades. The regrouping of a sequence or group of sequences to a different clade is therefore indicative of recombination between the study sequences and the respective bootstrap supported clade. Two TreeOrder Scans were generated in this study. The first scan (Figure 3.4) includes all human
genotypes of HBV A-G with all identified NHP variants. This scan confirms the effectiveness of the program in successfully identifying previously documented HBV recombinants. The second scan (Figure 3.5) includes only NHP to study recombination between species and sub-species of NHP. This scan allows for sub-species specific identification of recombination events taking place within study isolates.

3.2.6.2 Determination of recombination break points

In order to determine the specific recombination breakpoints of identified study recombinant ECO50003 Grouping Scan was carried out using SSE v1.1 (Simmonds, 2012). The query sequences were tested against sequences assigned to designated tag groups. Dependent up on the query sequence, assigned tag groups included; P.t.e, P.t.t, P.t.v, P.t.s, G. gorilla and HBV-C. The program then scored how deeply embedded the test sequences were within each of the assigned clades, generated by phylogenetic analysis of the pre-assigned tag groups. This study examined the breakpoints of previously identified recombinants FJ798098 P.t.t, FJ798099 P.t.e, AB498266 P.t.s, AB046525 P.t.t, study recombinant ECO50003 G. gorilla and identified non-recombinant AM117396. The successful correlation of the Grouping Scan results with those previously determined verified the effectiveness of the program in identifying recombination breakpoints and the origins of the recombinant regions.

3.3 Results

3.3.1 Isolation of HBV from wild caught non-human primates samples between August 2004 and August 2009, Cameroon

A total of 164 non-human primate plasma samples from 11 gorillas, 62 chimpanzees and 91 Old World Monkeys (OWM) were screened for the presence of HBV DNA. PCR screening with HBV screening primers S1/S5 and S3/S6 (MacDonald et al.,
2000) detected HBV DNA in 9/73 (12%) apes corresponding to 2/11 gorillas (18%) and 7/62 chimpanzees (11.3%) and 2/91 (2.2%) OWM. Complete HBV genomes were obtained from only the isolates of 2 Gorilla gorilla and 6 chimpanzees (4 P. t. elliottii, 2 P. t. troglodytes), while both the Old World Monkey isolates (1 Grey cheeked mangabe and 1 Mandrill) and 1 chimpanzees were positive only with the screening primers originally used. Samples ranged in age from 8 months to 9 years, sampled between 1996 and 2006. The likely wild origins of these animals are estimated based on the known natural habitats and available data from both the current holding sanctuary and the previous holders. Unfortunately at the time when samples were obtained sub-species of chimpanzees were not segregated but gorillas were isolated from chimpanzees in separate enclosures and were not in direct at any point of their rescue.

**TABLE 3.5: Specimen isolation data for all HBV positive NHPs.** Samples originated from Cameroon sanctuary, where sub-species of chimpanzees are co-housed together but segregated from gorillas. Prior to their arrival to the sanctuary they had been previously held at other locations within the country and wild origins based on data available to the sanctuary operators.

<table>
<thead>
<tr>
<th>Specimen No.</th>
<th>Mito.</th>
<th>HBV Variant</th>
<th>Location</th>
<th>Date of arrival</th>
<th>Age</th>
<th>Date of sampling</th>
<th>Previous Holder</th>
<th>Likely wild origin</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gorilla ECO50003</td>
<td>G.g</td>
<td>G.g</td>
<td>Limbe Wildlife Centre</td>
<td>13-9-03</td>
<td>9 years</td>
<td>19-8-04</td>
<td>Nkoma, Mvang, South Region</td>
<td>South Region</td>
</tr>
<tr>
<td>Gorilla ECO50063</td>
<td>G.g</td>
<td>G.g</td>
<td>Limbe Wildlife Centre</td>
<td>17-12-96</td>
<td>9 months</td>
<td>22-6-05</td>
<td>Bertoua, East Region</td>
<td>East Region</td>
</tr>
<tr>
<td>Chimpanzee ECO50083</td>
<td>P.t.e</td>
<td>P.t.e</td>
<td>Limbe Wildlife Centre</td>
<td>4-10-05</td>
<td>1.5 years</td>
<td>12-10-05</td>
<td>Garoua Zoo North Region</td>
<td>Adamaoua or East Region</td>
</tr>
<tr>
<td>Chimpanzee ECO50210</td>
<td>P.t.e</td>
<td>P.t.e</td>
<td>Limbe Wildlife Centre</td>
<td>31-12-04</td>
<td>8 months</td>
<td>05-12-06</td>
<td>Bachou, Manyu, South Region</td>
<td>Banyang Mbo Wildlife Sanctuary</td>
</tr>
</tbody>
</table>
3.3.2 Determination by phylogenetic analysis of relationship between NHP HBV isolates

Phylogenetic analysis of the HBV strains using 415 bp S gene fragments confirmed the grouping of the novel chimpanzee HBV strains with previously published HBV chimpanzee sequences and the grouping of the two novel gorilla HBV sequences with previously published HBV strains AJ131657 and FJ98095-97 (Mugisha et al., 2011, Njouom et al., 2010). (The novel HBV sequence ECO50065LIP3J and FJ98095 were retrospectively identified as originating from the same gorilla in Limbe Wildlife Centre).

Complete genome sequencing of the eight study isolates as outlined previously (Figure 3.1) produced sequences of 3182-bp in length, comparable to reference chimpanzee and gorilla strains (Figure 3.1). Phylogenetic analysis based on the complete genome (Figure 3.2), demonstrated monophyletic groupings for each human genotype (A–H), a clade containing gibbon and orangutan variants and a third containing chimpanzee and gorilla HBV sequences, each supported by high bootstrap values (Figure 3.2). Sequences of all 6 novel chimpanzee HBV variants, clustered with HBV sequences previously obtained from P. t. ellioti regardless of their host sub-species (Table 3.3) (Njouom et al., 2010).

<table>
<thead>
<tr>
<th>Chimpanzee ECO51109</th>
<th>P. t. t</th>
<th>P. t. e</th>
<th>Mfou NP Sanctuary</th>
<th>25-5-00</th>
<th>1.25 years</th>
<th>08-08-06</th>
<th>Mfou, Central Region</th>
<th>Akom II, South Region</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chimpanzee ECO51212</td>
<td>P. t. t</td>
<td>P. t. e</td>
<td>Mfou NP Sanctuary</td>
<td>8-8-06</td>
<td>1 year</td>
<td>10-3-08</td>
<td>Unknown</td>
<td>East Region</td>
</tr>
<tr>
<td>Chimpanzee ECO51394</td>
<td>P. t. e</td>
<td>P. t. e</td>
<td>Mfou NP Sanctuary</td>
<td>13-9-05</td>
<td>1.5 years</td>
<td>26-8-09</td>
<td>Unknown</td>
<td>Centre Region</td>
</tr>
<tr>
<td>Chimpanzee ECO51377</td>
<td>P. t. e</td>
<td>P. t. e</td>
<td>Mfou NP Sanctuary</td>
<td>1-10-97</td>
<td>4.6 years</td>
<td>26-7-09</td>
<td>Unknown</td>
<td>Unknown</td>
</tr>
</tbody>
</table>

Species association and HBV recombination in NHPs
FIGURE 3.2: Phylogenetic analysis based on the complete HBV genome. Phylograms displaying phylogenetic trees based on the complete HBV genome; with HBV reference sequences from human genotypes (A-H). Relative species and sub-species HBV variants are identified as follows Pan troglodytes troglodytes▲, Pan troglodytes elliottii▼, Pan troglodytes verus■, Pan troglodytes schweinfurthii◆, Gorilla gorilla○ and Hylobates spp. □, and the host specific cluster is identified by
Species association and HBV recombination in NHPs

The trees were rooted with the woolly monkey HBV sequence, NC_001896. Sequences from this study are in bold, while recombinant HBV variants are underlined.

3.3.3 Mitochondrial sequencing for determination of NHP species and sub-species

In order to determine the species and sub-species of NHP associated with each HBV isolate from the study, mitochondrial sequencing was applied. Based on the amplified sequences it was possible to discriminate not only between species of NHP but also between sub-species of chimpanzee; P. t. troglodytes, P. t. ellioti, P. t. verus and P. t. schweinfurthii. Mitochondrial sequencing confirmed that the ECO50083LIP5, ECO50210LIP4, ECO51394CWP1.4, ECO51377CWP2 and ECO51109CWP4, ECO51212CWP6 HBV variants originated from chimpanzee subspecies P. t. ellioti and P. t. troglodytes respectively, while ECO50003LIP3 and ECO50065LIP3 were identified as gorilla-derived (Figure 3.3), suggesting that two study P. t. troglodytes carry P. t. ellioti HBV variants (Figure 3.3), as suggested by phylogenetic analysis clustering of all chimpanzee HBV isolates of this study within the P. t. ellioti clade (Figure 3.2). All study samples were obtained from captive settings where P. t. troglodytes and P. t. ellioti are frequently co-housed. It remains unclear whether these findings reflect the HBV genotype distribution among wild chimpanzees in Cameroon or whether infection of troglodytes by ellioti-derived strains occurred in captivity.
Species association and HBV recombination in NHPs

3.3.4 Recombination detection by Tree order scan analysis across HBV genome

Phylogenetic trees of 200 bp fragments incrementing by 50 bp across the entire genome were constructed to identify changes in phylogeny potentially indicative of recombination events. This procedure was automated by the program TreeOrder Scan in SSE v1.0 (Simmonds, 2012). The tree position of each sequence across the genome is recorded (y-axis) and colour-coded by HBV genotype and species (Figure 3.2) and by sub-species (Figure 3.3). Changes in the tree order of individual sequences or genotypes with 70% or greater bootstrap support are indicative of alterations in the phylogenetic relationships of clades and identify potential recombination breakpoints. Fragments sizes ranging from 100bp to 750bp with
increments between 20bp and 250bp were also tested to detect the occurrence of recombination events. The most effective parameters were determined to be 200bp fragments with a 50bp overlap that successfully detected all previously reported recombinants in addition to those discovered in the study detailed here.

Consistent with previous findings (Simmonds and Midgley, 2005), phylogenetic relationships between human genotypes changed between genome regions, leading to alterations in the branching order (Figure 3.4). For example, genotypes D and E were largely phylogenetically distinct across the genome, but between position 1950 and 2500 (the core gene); genotype E falls within the genotype D clade. Excluding gorilla and chimpanzee sequences, recombination events typically occur around positions 750, 900, 1600, 1950, 2500, 2650 and 2750, frequently coinciding with gene boundaries as previously described (Simmonds and Midgley, 2005).
Species association and HBV recombination in NHPs

**FIGURE 3.4: Tree Order Scan of NHP HBV sequences with reference sequences from all human HBV genotypes.** TreeOrder Scan of HBV sequences, indicating positions of individual sequences (y axis) in Phylogenetic trees generated from sequential 250-base sequence fragments, incrementing by 50 bases. Changes in sequence order as a result of changes in phylogeny at the 70% bootstrap level are shown. Sequences are colour coded by genotype and host species, as indicated by the labels in left and right margin: genotype A, purple; B, light blue; C, wine; D, emerald; E, royal blue; F, orange; G, pale green; H, navy; Gorilla, blue (Gor); Chimpanzee, green (Pan); and Woolly monkey (WM-outgroup on line 1), red. For comparison the Tree Order Scan has been aligned with scale genome of HBV (top panel). Recombinant sequences are highlighted as by dashed lines; black gorilla/P.t.e ECO50003LIP3, green FJ798099 P.t.e/P.t.t, pink FJ798098 P.t.e/P.t.t, orange AB046525 P.t.t and purple AF498266 P.t.s.

In order to detect recombination events between HBV variants from different ape taxa, a Tree Order Scan was performed with all HBV reference and study sequences from gorillas and chimpanzees with a Hylobates pileatus sequences as an out-group (Figure 3.5). Outside of the core gene region, gorilla-derived variants were phylogenetically distinct from other ape-associated and human genotypes. However,
between positions 1560 and 2120 (in the X and pre/core gene), the gorilla isolate ECO50003LIP3 sequence grouped within the chimpanzee clade, indicative of a recombination event. ECO50003LIP3 showed 99.45% (99.0–99.9%) similarity with other gorilla variants between positions 1 and 1559 and 99.65% (99.5–99.8%) similarity between positions 2121–3272, and an average 95.5% and 97.4% similarity with P.t.ellioti variants across the same respective regions.

This analysis also identified recombination in the sequence A498266 (Figure 3.5), a P.t.schweinfurthii HBV isolate previously identified as a recombinant between human genotype C and chHBV (Magiorkinis et al., 2005). For this sequence, a 500 nt region between positions 550–1050 nt grouped with species C while the remainder of the genome grouped with P.t.troglodytes sequences, consistent with previous findings (Magiorkinis et al., 2005). The recently described HBV sequence from P.t.schweinfurthii (Mugisha et al., 2011) may therefore represent the “original” P.t.schweinfurthii sequence, from which the recombinant arose (Mugisha et al., 2011). The inclusion of these sequenced isolates described in our investigation supports the formation of a P.t.schweinfurthii species-specific clade (Figure 3.2) that includes the genotype C/chHBV recombinant. Corroborating evidence of recombination was identified in Cameroon chimpanzee sequences, FJ09898.1 and FJ09899.1; between positions 820 and 1300 nt, traversing a partial region of the polymerase gene confirming previous findings on recombinant P.t.t/P.t.e HBV variants (Njouom et al., 2010).
FIGURE 3.5: Tree Order Scan analysis study HBV isolates with available NHP HBV sequences. Tree Order Scan of HBV sequences, indicating positions of individual sequences (y axis) in phylogenetic trees generated from sequential 250-base sequence fragments, incrementing by 50 bases. Changes in sequence order as a result of changes in phylogeny at the 70% bootstrap level are shown. Sequences are colour coded by host species and sub-species of chimpanzee, as indicated by the labels in left and right margin: Gorilla gorilla, blue (Gor); Pan troglodytes troglodytes, yellow (Ptt); Pan troglodytes elliottii, green (Pte); Pan troglodytes verus, purple (Ptv); Pan troglodytes schweinfurthii, violet (Pts); and Hylobates pileatus (Hyl) (out-group-line 1-GII), red. For comparison the Tree Order Scan has been aligned with scale genome of HBV (top panel). Recombinant sequences are highlighted as by dashed lines; black gorilla/P.t.e ECO50003LI293, green FJ798099 P.t.e/P.t.t, brown FJ798098 P.t.e/P.t.t, orange AB046525 P.t.t and blue AF498266 P.t.s.

To confirm the TreeOrder scan and Grouping Scan analysis for study recombinant ECO50003, individual phylogenetic trees of the of pre-recombinant (Figure 3.6A), recombinant (Figure 3.6B) and post recombinant regions were generated (Figure 3.6C). Phylogenetic trees confirm the shift in the phylogenetic relationship of
recombinant *Gorilla gorilla* ECO50003 from the *G. gorilla* clade to the *P. t. ellioti* clade and back to the *G. gorilla* clade. It further identifies the shift in the phylogenetic relationships of previously identified recombinants FJ798098, FJ798099 and AB046525.

**FIGURE 3.6:** Phylogenetic analysis based on the recombinant region 1560-2120bp of HBV genome. Phylograms displaying phylogenetic trees based on (B) HBV recombinant region 1560-2120bp and equivalent fragments immediately preceding; (A) 999-1559bp and succeeding; (C) 2121-2681bp this region; with HBV reference sequences from human genotypes (A-H) and *Hyllobates* spp. collapsed and cropped. Relative species and sub-species of HBV variants are identified as follows *Pan troglodytes troglodytes* ▲ (*P.t.t*), *Pan troglodytes ellioti* ▼ (*P.t.e*), *Pan troglodytes verus* ■ (*P.t.v*), *Pan troglodytes schweinfurthii* ◦ (*P.t.s*), and *Gorilla gorilla* ○ (*G.g*), and the host specific cluster is identified by ]. The trees were rooted with the woolly monkey HBV sequence, NC_001896. Sequences from this study are in bold and identified recombinant ECO50003 and its shift in phylogeny is marked in red, while recombinant HBV variants previously reported have been underlined in green and where relevant their shift in phylogeny is identified by a green line.
3.3.5 Determination of recombination breakpoints in HBV genome of NHP study isolates

To confirm the position and phylogenetic grouping of the putative recombinant sequences identified by the TreeOrder scan, a Grouping Scan was performed (Simmonds, 2012). This examines how deeply embedded the test sequence is within clades formed by non-recombinant control sequences assigned into species-associated groups (P. t. e, P. t. t, P. t. v and Gorilla gorilla) (Figure 3.4). This method identified two changes in grouping of the query sequence, ECO50003LIP3 (Figure 3.5A) at position 1560 where it changed grouping from the gorilla HBV clade to the P. t. ellioti sub-species clade and a reversion to the gorilla clade at position 2120.

Grouping scan analysis of recombinant sequences FJ98098.1 and FJ98099.1 (Figure 3.6) and A498266 (Figure 3.7A) provides substantial support for the formation of recombinant regions between positions 550–1050 nt and 820–1300 nt respectively (Figure 3.6A). However, the Tree Order or grouping scan (Figure 3.3 and Figure 3.5B) methods provided no evidence for recombination in the P. t. troglodytes derived sequence, AM117396 (based on its grouping in Figure 3.1A) between chimpanzee sub-species.

Sequence AB046525 from P. t. troglodytes in Central Africa grouped separately from other P. t. t variants in core gene region (Figure 3.7B), consistent with past recombination with a divergent and currently uncharacterised genotype of HBV (Takahashi et al., 2000). Both TreeOrder scan analysis (Figure 3.3) and Grouping Scan analysis (Figure 3.7B) confirmed the rest of the genome groups consistently with P. t. troglodytes, while adopting an outlier position to all other chimpanzee and gorilla isolates in the core region.
Species association and HBV recombination in NHPs

**FIGURE 3.7:** Grouping Scan analysis of (A) Gorilla HBV study recombinant compared with (B) non-recombinant chimpanzee AM117396. Sequence fragments of 250 bases incrementing by 100 bases with 100 bootstrap replicates, were used to compare and analyse breakpoints present in study HBV isolates (A) study recombinant *Gorilla gorilla* HBV sequence (ECO50003); (B) *P.t.troglodytes* AM117396 to sequence groups from *Gorilla gorilla* (red), *Pan troglodytes ellioti* (blue), *P.troglodytes* (green), *P.t.versus* (yellow), *P.troglodytes* and *P.t.schweinfurthii* (purple) and human genotype HBV/C (light blue) with respect to A498266. Values >0.5 indicate clustering within the indicated group.
FIGURE 3.8: Grouping Scan analysis of HBV recombinants from P.t.elliottii and P.t.troglodytes identified in this study. Sequence fragments of 250 bases incrementing by 100 bases with 100 bootstrap replicates, were used to compare and analyse breakpoints present in study HBV isolates (A) P.t.troglodytes/P.t.elliottii recombinant FJ98098.1 (B) P.t.elliottii/P.t.troglodytes recombinant FJ98099.1, to sequence groups from Gorilla gorilla (red), P.t.elliottii (blue), P.t.troglodytes (green), P.t.verus (yellow), P.t.schweinfurthii (purple) and human genotype HBV/C (light blue) with respect to A498266. Values >0.5 indicate clustering within the indicated group.
FIGURE 3.9: Grouping scan analysis of HBV recombinants AB498266 and AB046525 from *P.t.schweinfurthii* and *P.t.troglodytes* respectively. Sequence fragments of 250 bases incrementing by 100 bases with 100 bootstrap replicates, were used to compare and analyse breakpoints present in study HBV isolates (A) *P.t.schweinfurthii* isolate A498266; (B) *P.t.troglodytes* recombinant AB046525, to sequence groups from *Gorilla gorilla* (red), *P.t.elliotti* (blue), *P.t.troglodytes* (green), *P.t.verus* (yellow), *P.t.schweinfurthii* (purple) and human genotype HBV/C (light blue) with respect to A498266. Values >0.5 indicate clustering within the indicated group.
3.4 Discussion

3.4.1 Screening of NHPs for presence of human and species-specific variants

In this study 164 plasma samples from great apes and monkeys from Cameroon were screened for the presence of HBV-DNA (Table 3.1). The prevalence amongst chimpanzees in our study was found to be 9.7% (6/62), and 18% (2/11) in gorillas. This confirms previous findings on the existence of HBV in great apes in the wild (Makuwa et al., 2005, Makuwa et al., 2006, Njouom et al., 2010, Simmonds and Midgley, 2005) and the rates are comparable to those observed in human populations in areas of endemic infection, such as Central Africa and South East Asia. Prevalence observed in this study is within the range of recent studies where the prevalence of active HBV infection (DNA-positive in plasma) was 15% (8/53) and 8% (3/26) in gorillas and 18% (40/205) and 5% (8/156) in chimpanzees (Njouom et al., 2010, Starkman et al., 2003, Sa-Nguanmoo et al., 2009). The frequency of infection detected amongst study primates is also comparable to that recently detected within bat sub-species from Africa and Central America (Drexler et al., 2013b). HBV related viruses were detected in 9.3% (5/54) New World bats (Uroderma bilobatum), and in 7.9% (4/51) and 6.3% (1/16) of Old World bats from Gabon (Hipposideros cf. ruber and Rhinolophus alcyone).

3.4.2 Complete genome analysis of study sequences isolates from NHPs

Eight new complete HBV genomes were obtained in the current study from two gorillas and six chimpanzees born in the wild (Figure 3.10) and subsequently held in rescue sanctuaries. Gorilla sequence ECO50065LIP3 was almost identical to the previously described sequence, FJ798095 (Njouom et al., 2010), and retrospective analysis revealed that these sequences originated from the same animal in Limbe Wildlife Centre.
FIGURE 3.10: Geographical locations of suggested wild capture of study isolates in Cameroon. The provincial map of Cameroon provides geographical data on the possible wild origins of study samples taken from non-human primates. Where specific information is available for wild origin the sites are denoted with red arrow of approximate geographical location.

The phylogenetic tree comparing the six complete chimpanzee HBV sequences with other previously reported HBV isolates from NHPs grouped all with previously identified P.t.elliottii variants (Njouom et al., 2010) (Figure 3.1). Mitochondrial sequencing confirmed that of the six complete chimpanzee genomes grouping in the P.t.elliottii clade, two of these were recovered from P.t.troglodytes and the remaining four from P.t.elliottii (Figure 3.2). Current analyses cannot determine whether these two cross species infections occurred in the wild or through contact with infected P.t.elliottii chimpanzees while in captivity although the latter is highly likely given the mixing of chimpanzee subspecies in sanctuaries.

Phylogenetic analysis showed the grouping of HBV sequences from great apes in Asia (gibbons and orangutans) and great Apes in Africa (gorillas and chimpanzees).
into two distinct clusters, one containing *Hylobates* spp. and one contain all Pan subspecies and gorillas (Figure 3.2). This finding is congruent with previous studies suggesting geographical association of HBV in NHP (Makuwa *et al.*, 2007, Starkman *et al.*, 2003, Njouom *et al.*, 2010). Within the African clade, all HBV variants infecting chimpanzees fell into distinct sub-species specific clades in addition to a single gorilla clade (Figure 3.2). Phylogenetic analysis based on the complete genome found significant bootstrap support for the formation of four HBV clusters that, excluding likely cross-species transmissions, corresponded with *P. troglodytes* subspecies: *P.t.troglodytes* (89% bootstrap), *P.t.verus* (100% bootstrap), *P.t.ellioti* (100% bootstrap) and *P.t.schweinfurthii* (100% bootstrap). The *P.t.ellioti* and *P.t.troglodytes* formed species-specific clades despite the fact that both chimpanzee sub-species can share habitats in the wild (Figure 3.11), confirming previous studies on HBV variants in chimpanzees in Cameroon (Njouom *et al.*, 2010).

The existence of *P.t.troglodytes* and *P.t.ellioti* associated variants of HBV, as is the case for other chimpanzee subspecies (*P.t.verus* and *P.t.schweinfurthii*) requires further investigation of variants infecting chimpanzees in the wild in Cameroon, in particular in regions where these sub-species may converge (Figure 3.8), for example around the confluence of the Mbam and Sanaga Rivers (Gonder *et al.*, 2011). The observation of cross-species infections and recombination events for HBV infections also provides an additional reason for ensuring that captive chimpanzees are correctly identified to subspecies and segregated appropriately to avoid the creation of recombinant HBV variants with potentially different pathogenicities and transmission patterns.
Detection of HBV recombination between *Gorilla gorilla*/ *P. t. troglodytes*

The co-infection of one individual or animal with two different HBV genotypes can result in the exchange of genetic material between the two viral strains. Recent research has reviewed and examined the presence of HBV recombination using the TreeOrder scan method applied in our study (Simmonds and Midgley, 2005, Yang *et al.*, 2006). Almost all recombinants detected (90%) were between B/C or A/D in addition to other recombinants between A/E, A/G, C/D, C/F, and B/C/U where U is an unknown. The predominant recombination breakpoints were found to be concentrated around X gene region (1640-1900nt), preS1/S2 regions (3150-100nt), the 3’ of the surface gene (650-830nt) and the 3’ end of the core gene (2330-2450nt) (Simmonds and Midgley, 2005).
Recent phylogenetic analysis of the *P. t. schweinfurthii* HBV strain; confirmed by our GroupScan analysis; (Figure 3.9A) showed further evidence of interspecies recombination between HBV infecting chimpanzees and the human HBV-C genotype strain (Hu *et al.*, 2000, Magiorkinis *et al.*, 2005). Phylogenetic trees of the recombinant region and equivalent fragments either side, inclusive of all reference and study sequences, confirm the sub-species association of HBV in NHPs (Figure 3.4 A to Figure 3.4 C). The phylograms also support the recombinant data of the Tree order and Grouping scan analysis, with respect to the location and confidence level for the recombinant region and sequence (Figure 3.4 A).

Recombination affecting a short region between either end of the polymerase gene (partial X gene and Pre-core/core) in one *Gorilla gorilla* isolate is the first recorded occurrence of recombination between chimpanzee- and gorilla-derived HBV variants. The recombination event between the gorilla and *P.t.ellioti* variants likely occurred in the wild as gorillas and chimpanzees are never co-housed in captivity. The position of the breakpoint region is close to several documented previously in human genotypes (Simmonds and Midgley, 2005) (Figure 3.3). A recent study of duck HBV (Liu *et al.*, 2010b) recorded a similar breakpoint event between position 1010–2304 bp, incorporating the region of the X gene, which is believed to promote cell growth and inactivate growth regulating molecules (Martin *et al.*, 2011, Li *et al.*, 2010). The correlation of HBV sequences with the different subspecies of chimpanzees indicates either that the HBV strains and their hosts have co-evolved or alternatively have diverged through allopatric separation.

### 3.4.4 Co-divergence and allopatric speciation as models of HBV evolution

The co-divergence hypothesis for the distribution of non-human HBV genotypes in Africa and South East Asia presupposes that the distinct variants of HBV found in different ape species and subspecies arose during the period of their evolution, over
the past 5-7 million years (MacDonald et al., 2000). While the different geographical sub-species of chimpanzees have HBV sequences that constitute distinct phylogenetic clades and would support this hypothesis, isolates from different sub-species in the same geographical areas have also been shown to carry sub-species specific HBV variants (Njouom et al., 2010). Such a hypothesis implies an extraordinarily slow maintained substitution rate of HBV; the 5% divergence between gorilla and chimpanzee variants requires a minimum substitution rate of $3-5 \times 10^{-9}$ substitutions per site per year (SSY), far lower than mammalian coding region substitution rates and quite distinct from the $10^{-4}$–$10^{-5}$ SSY rates estimated for HBV over shorter periods (Hannoun et al., 2000a). The co-divergence hypothesis would additionally predict that HBV variants infecting gorillas should be more divergent from and take an outlier phylogenetic position to the subspecies-associated variants of chimpanzees that would have diverged from each other between 0.8–1.5 million years ago (Starkman et al., 2003).

This is clearly not the case in the phylogenetic analysis shown in Figure 3.2 and Figure 3.6, where the gorilla clade adopts an internal branching position among chimpanzee sub-species associated variants. Similarly, the co-divergence hypothesis cannot explain the inlier position of orangutan derived HBV variants within the gibbon clade (Sall et al., 2005), nor the substantial sequence diversity in HBV variants infecting humans and their outlier position to NHP-derived HBV sequences. This model would require that HBV genotypes infecting humans arose through multiple cross-species transmission events with primates carrying various species-specific genotypes. A similar scenario resulted in the origins of HIV-1 and HIV-2, where HIV-1 resulted from at least 3 separate cross-species transmission events between chimpanzees and humans (Gao et al., 1999) and HIV-2 infection arose through cross-species transmission from sooty mangabeys in West Africa (Feng et al., 1992). A primate origin of HBV is further supported by the fact that the areas of highest human HBV prevalence are areas in which contact with NHPs and cross-
species transmission is more probable, sub-Saharan Africa, South East Asia, and South America. However, no HBV genotypes have been detected as shared between primates and humans occupying same geographical areas in this or previous studies, with the exception of a genotype E variant detected in a chimpanzee (Takahashi et al., 2000), indicating cross-species transmission has not occurred.

On the other hand, co-divergence of virus and host and the implied extremely low long-term viral substitution rates have been observed within other primate associated viruses, including simian immunodeficiency virus (SIV) and simian foamy virus (Sa-Nguanmoo et al., 2009, Starkman et al., 2003). In the case of HBV, substitutions may accumulate slowly as a result of the extreme constraints on sequence change in the HBV genome imposed by the extensive use of overlapping reading frames for protein coding, as well as RNA secondary structures required for genome transcription and translation (Simmonds and Midgley, 2005, Starkman et al., 2003). However, co-divergence does not explain how HBV-like viruses infecting rodents, squirrels and birds could have become so divergent from human and primate variants over period perhaps only 10–20 times as long as the period of ape divergence.

The alternative hypothesis, which could account for the pattern of sequence diversity of HBV in NHPs, is divergence through allopatric (geographical) separation but with ongoing transmission of HBV between ape species and subspecies. This alternative hypothesis would also account for the internal branching position of the gorilla clade within the chimpanzee derived HBV sequences. As previously discussed, it also accounts for the inlier position of orangutan-derived sequences deep within the gibbon clade, and their close genetic relationship with the sympatric Hylobates agilis gibbon species (Sall et al., 2005, Noppornpanth et al., 2003). These two species occupy proximal or overlapping habitats in Borneo, while HBV variants infecting gibbons from elsewhere in Asia group separately (Starkman et al., 2003, Warren et al., 1999). The lack of sequence diversity between infected orangutans
implies a more recent introduction of HBV into this species (Warren et al., 1999). Efforts to identify recombinants might therefore focus on geographical regions where different non-human primate species and sub-species come into contact, for example the upper reaches of the Sanaga River in Cameroon where *P. t. elliotti* and *P. t. troglodytes* may occasionally mix (Gonder et al., 2011) and southern Cameroon where *P. t. troglodytes* and *G. gorilla* distributions overlap (Figure 3.10). The later hypothesis may potentially explain why no recombinant variants have been detected in *P. t. verus*, a subspecies found in West Africa that is geographically isolated from other non-human ape species. The timescale for the proposed geographical isolation of HBV variants infecting different NHP species and sub-species is not known. If we take the substitution rate for HBV measured over short periods (Zhou and Holmes, 2007), the introduction and geographical differentiation of HBV in African apes may have occurred relatively recently indeed, approximately 5,000 years using the previously described substitution rate of $10^{-5}$ substitutions per site per year (Fares and Holmes, 2002, Hannoun et al., 2000a).

### 3.4.5 Emergence of potential reservoir for primate hepadnaviruses

Recent discovery of a novel hepadnavirus infecting 3 species of bats in Africa and Central America sheds new light on the potential origins HBV infections. The 3 viral species detected TBHBV/RBHBV/HBHBV had nucleotide divergence of at least 35% from all other known hepadnaviruses and varied from 3,149 to 3,377 nt in length with genome organisation comparable to other orthohepadnaviruses (Drexler et al., 2013b). Phylogenetic analysis of the complete genome with other members of the *Hepadnaviridae* identified two distinct monophyletic clades comprising the Old and New Worlds bat viruses, respectively. Furthermore the bat variants formed a basal position in the tree to all primate associated HBV variants. The possibility of bats as the origins for primate hepadnaviruses has again raised speculation on their evolution. Bats have been shown to be major reservoirs for other human pathogens including paramyxoviruses, lyssaviruses, and sever acute respiratory syndrome
(SARS)-related coronaviruses, (Chua et al., 2000, Li et al., 2005, Lau et al., 2005, Calisher et al., 2006). Bats infected with species-specific variants of HBV displayed high viral loads and hepatic tropism was confirmed. Furthermore the recent study determined successful infection of human hepatocytes with the TBHBV variant (Drexler et al., 2013b). Such findings may have significant implications for the future management and eradication of HBV globally. This may be particularly important for those people living in regions where bush meat from bats in hunted and consumed regularly and therefore the risks for cross-species transmission or potential recombination generating new infectious variants is increased.

3.4.6 Conclusions

Complex epidemiological factors such as transmission routes affect the pool of circulating HBV variants, although their spread may be enhanced by the evolution of recombinant variants, allowing the virus to transmit more efficiently between species. Phylogenetic studies have previously indicated that recombination events in HBV are quite common (Sugauchi et al., 2002b, Simmonds and Midgley, 2005) and recombinant strains have been shown to possess distinct biological features and produce different clinical outcomes compared to their parental strains (Liu et al., 2010a). Further work is required to investigate the distribution of HBV recombinants in Cameroon, their potential impact on host species and the evolution of HBV in NHPs in the wild.

The evidence of animal reservoirs, cross-species transmission and recombination between human and ape HBV variants has important implications for the eradication of HBV worldwide. The transmission of recombinant gorilla/chimpanzee HBV and endemic infection among apes in the wild will hinder efforts to eradicate HBV globally, particularly in regions where apes and humans come into contact (MacDonald et al., 2000). Further sampling of isolated populations of NHPs, humans and bats in areas of sympatry is required to further investigate the currently
conflicting evolutionary hypotheses for HBV diversity and to determine how and when HBV spread between Africa and Asia. Understanding the relationship between human and NHP HBV variants and the novel bat HBV isolates will aid in resolving this question and explain the possible role each of these species has played in the origin and dissemination of HBV worldwide. Future research screening and clinically examining bat populations in areas of high HBV infections in Africa and South East Asia is required to elucidate the origins of primate HBV. This research will also aid our understanding of the potential threat if any, posed to global health by the existence of a bat reservoir for hepadnaviruses, given their close association with other pathogenic human viruses.
Chapter 4
Non-primate hepacivirus; detection and characterisation in domestic horses in the United Kingdom

4.1 Introduction

4.1.1 Pandemic spread of HCV in the Western world.

Developing countries and the Western world saw the dramatic spread of HCV during the twentieth century, sparked by the ready availability of parenteral routes of transmission associated with blood transfusions, vaccinations, other medical procedures and more recently injecting drug use. The discovery of HCV in the late 1980s was driven by the concerns of clinicians and epidemiologists who increasingly witnessed cases of non-A, non-B hepatitis associated with blood transfusions and medical procedures using plasma-derived blood products and unsterilized needles (Prince et al., 1974, Feinstone et al., 1975, Alter et al., 1975). Development of effective diagnostic methods made evident the extent of HCV dissemination worldwide. Currently an estimated 170 million are chronically infected globally, representing 3% of the world’s population, resulting in an extensive healthcare burden from chronic liver disease and cirrhosis (Pawlotsky, 2003, Hoofnagle, 2002, Seeff, 2002). Modern transmission routes associated with injecting drug use and needle sharing is the primary route of viral transmission following the introduction in the 1990s of effective blood donor screening programs and steps to inactivate blood products (Nelson et al., 2011).

The hypothesis for the relatively recent dissemination of HCV into the western world can only be indirectly inferred, due to a lack of sample availability pre World War II. Modelling based on the evolutionary history of currently circulating variants of the virus is the best available method for indirectly inferring timescale for viral spread.
HCV transmission is inefficient and rare through sexual contact or from mother to child (Wasley and Alter, 2000, Pradat and Trepo, 2000, Thomas, 2000), therefore a timescale for the spread of HCV can be inferred by examining development and availability of primary parenteral transmission routes. The use of unscreened blood in transfusions, unsterilized needles for vaccinations and in modern times, exposure to contaminated needles associated with IDU points to a recent timeframe for the spread of HCV. Such practices were not commonplace prior to World War II supporting the hypothesis for the spread of genotypes 1b and type 2 subtypes during the 1940s-50s and in the 1960s the transmission among IDUs (Pybus et al., 2001, Cochrane et al., 2002).

HCV subtypes are epidemiologically distinct, targeting different risk groups and geographical regions that reflect their recent epidemic spread. Genotypes 3a and 1a largely infect IDUs in Northern Europe; while 4a is frequently detected in the Middle East and particularly Egypt, further evidence for the recent dissemination of HCV into these populations. (Van Asten et al., 2004, Kamel et al., 1994, Abdel-Aziz et al., 2000, Ray et al., 2000). One hypothesis for the 10-20% infection rates observed in Egypt is that extensive nationwide transmission occurred after a countrywide public health campaign inoculating against Schistosoma infections (Arthur et al., 1995). The effect of such measures is evident in the sharp decline in HCV infections in persons born after the discontinuation of the campaign in the 1970s (Frank et al., 2000, El-Zayadi et al., 1997, Angelico et al., 1997, Pybus et al., 2003, Kamel et al., 1994). The Egypt case study provides further supportive evidence for the recent emergence of HCV into developing countries through the utilisation of modern transmission routes.

4.1.2 HCV genetic diversity and endemic circulation

HCV shows considerable genetic diversity; 7 genotypes with >30% nt sequence divergence from each other (Simmonds et al., 2005). Until recently, GBV-B was the
only other virus classified as a member of the \textit{Hepadivirus} genus, recovered from a laboratory-infected tamarin (Simons \textit{et al.}, 1995b, Muerhoff \textit{et al.}, 1995). To date only a single isolate of GBV-B has been identified and the ultimate origin remains unidentified.

Of the 7 HCV genotypes several are associated with suspected endemic source areas in central and western sub-Saharan Africa (genotypes 1, 2, and 4) (Ruggieri \textit{et al.}, 1996, Jeannel \textit{et al.}, 1998, Candotti \textit{et al.}, 2003, Ndjomou \textit{et al.}, 2003) and South East Asia (genotypes 3 and 6) (Smith \textit{et al.}, 1997, Tokita \textit{et al.}, 1994a). Additionally, extensive neighbouring regions of the African continent are dominated by individual genotypes. Western Africa is predominately genotype 2 (Candotti \textit{et al.}, 2003, Jeannel \textit{et al.}, 1998, Mellor \textit{et al.}, 1995, Ruggieri \textit{et al.}, 1996). Central African countries of Congo, Cameroon and Gabon are genotypes 1 and 4 (Bukh and Miller, 1994, Fretz \textit{et al.}, 1995, Menendez \textit{et al.}, 1999, Xu \textit{et al.}, 1994, Li \textit{et al.}, 2009, Li \textit{et al.}, 2012), and limited available studies indicate genotypes 5 and 7 are largely the source of HCV infections in Central and Southern Africa (Murphy \textit{et al.}, 2007) (Smuts and Kannemeyer, 1995).

These regions harbour the greatest diversity of HCV subtypes, implying a long-term, endemic circulation of the virus over several hundred years. The spread of certain genotype variants from these populations, such as 1a and 1b to Western countries, 3a among injection drug users in Europe, and 4a to Egypt, where it was extensively transmitted by medical injections (Pybus \textit{et al.}, 2003, Frank \textit{et al.}, 2000, Magiorkinis \textit{et al.}, 2009, Pybus \textit{et al.}, 2005), show several parallels with the emergence and rapid spread of HIV-1 among new risk groups from a central African reservoir over a similar time frame (Gao \textit{et al.}, 1999).

\subsection*{4.1.3 Characterisation of HCV homologues in other mammals}

As previously discussed in section 1.4.4 homologs for HCV have now been detected in a range of mammals including rodents (Drexler \textit{et al.}, 2013a, Kapoor \textit{et al.},
2013b), bats (Quan et al., 2013), horses (Burbelo et al., 2012b, Lyons et al., 2012a), dogs (Kapoor et al., 2011) and more recently in Old World monkeys in Uganda (Lauck et al., 2013). The detection of these viruses has significantly broadened the focus of research from NHPs to include a much broader range of mammals.

Currently very little is known about the course of infection associated with these viruses and whether or not they are hepatotropic. However, unpublished data (Kapoor et al.) from current studies suggest that rodent hepacivirus (RHV) infects the liver and replicates in hepatocytes. At present there is minimal data relating to the clinical manifestation and disease course associated with NPHV infection. Results presented within this chapter seek to investigate the species distribution of NPHV or homologs in a range of mammalian species. The genetic relatedness of the novel isolates was compared by phylogenetic analysis to previously published genomes. Furthermore the availability of repeat samples from an actively viraemic horse allowed for the first investigation into the clinical manifestations, disease course and epidemiology of NPHV infection and was compared with those of HCV.

4.2 Materials and Methods

4.2.1 Selection of samples

A total of 552 samples were screened for the presence of NPHV RNA. Samples collected were from horses, dogs, cats, mice, pigs, and donkeys. All horse, cat, and donkey samples were sourced from either excess diagnostic samples or previously archived study samples from the Royal (Dick) School of Veterinary Studies, University of Edinburgh, where the laboratory investigation for this study was performed. Buccal swab samples were obtained from dogs (Canis lupus familiaris) undergoing veterinary examination at the Edinburgh Dog and Cat home. Bronchoalveolar lavage samples (BAL) were collected from dogs undergoing investigation of respiratory disease.
Venus blood samples were collected from 353 nonprimates comprising 99 dogs, 158 horses and donkeys (142 Equus ferus caballus, 16 Equus africanus asinus), 56 cats (Felis catus), 63 rodents (47 Apodemus sylvaticus, 8 Mus musculus, 5 Myodes glareolus, 3 Microtus agrestis), and 40 pigs (Sus scrofa). Plasma was separated by centrifugation and frozen at −80°C until testing. Lung, liver, and spleen samples were obtained from dogs during necropsy at the pathology department of the Royal (Dick) School of Veterinary Studies and placed in RNAlater (QIAGEN, Crawley, UK) before RNA extraction. Samples of mouse liver were collected from all rodents in East Lothian, Scotland except 2 Mus musculus for which liver samples were unavailable. All clinical sampling was undertaken with full owner consent and in line with Royal (Dick) School of Veterinary Studies institutional and UK ethical guidelines.

4.2.2 CHV and NPHV Screening

4.2.2.1 CHV/NPHV RNA Transcript

Screening for CHV and NPHV infections was performed by PCR; serologic screening for antibodies against CHV/NPHV was precluded by the non-availability of NS3 antigen used in a previous study (Burbelo et al., 2012b). To validate the PCR, RNA transcripts were generated from a plasmid containing partial CHV NS3 cDNA by using the Ambion T7 transcription kit (Promega Corp., Southampton, UK). Transcripts were purified with the RNeasy kit (QIAGEN), and concentrations were determined by using the NanoDrop 2000 (NanoDrop Products, Wilmington, DE, USA).

4.2.2.2 Extraction of viral RNA

RNA extractions were performed on 140 μL of plasma or respiratory sample by using the QIAmp viral extraction kit (QIAGEN) according to the manufacturer's instructions and eluted in a final volume of 60 μL. All tissue samples were
homogenized in lysis buffer; RNA was extracted by using the RNeasy Mini Kit (QIAGEN) according to instructions and eluted in a final volume of 60 µL.

4.2.2.3 Isolation of peripheral blood mononuclear cells
Peripheral blood mononuclear cells were separated from whole blood immediately after collection by centrifugation on a Ficoll-Hypaque density gradient by using Histopaque 1077 (Chapter 2) according to manufacturer’s instructions (Sigma Aldrich, St. Louis, MO, USA) and RNA was extracted by using QIAmp RNA blood mini kit as instructed (QIAGEN) and eluted in final volume of 100 µL.

4.2.2.4 Generation of cDNA
RNA was converted to cDNA by using random hexamers with the Reverse Transcription System A3500 (Promega) and then used in nested PCR with previously published CHV NS3 primers (Kapoor et al., 2011) Chv-0F1, Chv-0R1S1, Chv-0F2, and Chv-0R2 and new equine-based NS3 primers (Table 1) and amplified by using 2 rounds of 30 cycles at 94°C for 18 s, 50°C for 21 s, and 72°C for 1.5 min; and 1 cycle of 72°C for 5 min, with 2 µL of first-round amplicon added to the second round. Degenerate equine- and canine-based NS3 primers were designed on the basis of the sequence variability observed in the published NPHV sequences and used to additionally screen all samples from dogs and equids. The CHV NS3 transcript was tested by using both NS3 primer sets and used as a control in screening, with sensitivity of 0.5–5 RNA copies in a reaction (Tables 4.1).
TABLE 4.1: Titration of NPHV-NS3 transcript tested with NS3 primers to determine PCR sensitivity for NPVH RNA in samples from domestic horses in the United Kingdom. Serial dilutions of NPHV transcript were tested with study-designed primers for NS3 and previously published primers sets for the detection of CHV.

<table>
<thead>
<tr>
<th>Transcript RNA copies/mL</th>
<th>Published NS3</th>
<th>New NS3</th>
</tr>
</thead>
<tbody>
<tr>
<td>5 x 10^6</td>
<td>2/2</td>
<td>2/2</td>
</tr>
<tr>
<td>5 x 10^5</td>
<td>2/2</td>
<td>2/2</td>
</tr>
<tr>
<td>5 x 10^4</td>
<td>2/2</td>
<td>2/2</td>
</tr>
<tr>
<td>5,000</td>
<td>2/2</td>
<td>2/2</td>
</tr>
<tr>
<td>500</td>
<td>2/2</td>
<td>2/2</td>
</tr>
<tr>
<td>50</td>
<td>2/2</td>
<td>2/2</td>
</tr>
<tr>
<td>5</td>
<td>2/2</td>
<td>2/2</td>
</tr>
<tr>
<td>0.5</td>
<td>0/2</td>
<td>0/2</td>
</tr>
<tr>
<td>0</td>
<td>0/2</td>
<td>0/2</td>
</tr>
</tbody>
</table>

4.2.2.5 PCR screening

To confirm positive results of screening, we designed degenerate primers derived from the NPHV sequences for the 5’-UTR and NS5B (Table 1). For all detected positive results, we used SuperScript III One-Step RT-PCR (Life Technologies, Paisley, UK) with 6 µL of RNA and cycling conditions as published (Burbelo et al., 2012b) with 1 of the following first-round primer sets: EQ5→UTROS and EQ5→UTROAS or EQNS5BIS and EQNS5BIAS (Table 4.2). From the first round, 2 µL was added to the second-round PCR with respective forward and reverse primer sets: EQ5→UTRIS and EQ5→UTRIAS or EQNSBIS2 and NS5BIAS2, with the following cycling conditions; 30 cycles at 94°C for 18 s, 50°C for 21 s, and 72°C for 1.5 min; and 1 cycle of 72°C for 5 min.

TABLE 4.2: Nonprimate hepaciviruses primer sequences for 5’UTR, NS3, and NS5B in samples from horses, United Kingdom*

<table>
<thead>
<tr>
<th>Primer</th>
<th>Position</th>
<th>Sequence, 5’→3’</th>
</tr>
</thead>
<tbody>
<tr>
<td>EQ5→UTROAS</td>
<td>Forward outer sense</td>
<td>ACA YYA CCA TGT GTC ACT CCC CCT</td>
</tr>
<tr>
<td>EQ5→UTROAS</td>
<td>Reverse outer antisense</td>
<td>CYC ATG TCC TAT GGT CTA CGA GA</td>
</tr>
<tr>
<td>EQ5→UTRIS</td>
<td>Forward inner sense</td>
<td>ACA CGG AAA YGG GTT AAC CAY ACY C</td>
</tr>
<tr>
<td>EQ5→UTRIS</td>
<td>Reverse inner antisense</td>
<td>GCC CTC GCA AGC ATC CTA TCA G</td>
</tr>
<tr>
<td>EQNS3OS</td>
<td>Forward outer sense</td>
<td>ATW TGT GAT GAR TGC CAY AGY AC</td>
</tr>
<tr>
<td>EQNS3OAS</td>
<td>Reverse outer antisense</td>
<td>TAG TAG GTB ACA GCR TTA GGY CC</td>
</tr>
<tr>
<td>EQNS3IS</td>
<td>Forward inner sense</td>
<td>TCY AAR GGT GTD AAG CTT GTT GT</td>
</tr>
</tbody>
</table>
**4.2.3 Sequencing of PCR amplicons**

Positive second-round PCR amplicons were sequenced in both directions by using the inner sense and inner antisense primers used in the second round of amplification. Sequencing was executed by using Big Dye Terminator version 3.1 (Applied Biosystems, Paisley, UK) according to the manufacturer's instructions. Sequences were analysed by using SSE version 1.0 software (Simmonds, 2012). Sequences obtained in this study have been assigned the GenBank accession nos. JX948116–JX948121.

**4.2.4 Determination of NPHV viral load**

Viral loads of positive samples were determined by real-time quantitative PCR and a standard calibration curve generated from a dilution series of the NS3 transcript. Dilutions were prepared from the CHV NS3 transcript from concentrations of $10^6$ to 1 copy/µL; 5 µL of transcript RNA was used to generate cDNA with random hexamers and reverse transcription. Five-microliter aliquots of cDNA were assayed in triplicate for the 3 positive samples in the same way. To quantify viral loads of positive samples, EQNS3IS and EQNS3IAS primers were used with 4 µL of cDNA in the SensiFAST SYBR Hi-ROX Kit (BioLine, London, UK) per manufacturer’s instructions (Chapter 2) with the exception that the annealing temperature was reduced to 50°C and the extension time extended to 15 s. Samples were analysed in triplicate and fluorescence measured by using the Rotor-Gene Q system (QIAGEN). Viral loads were read from the standard curve generated and converted to RNA copies/mL for sample volume used in extraction and elution of the RNA.
4.2.5 Sequence alignment and phylogenetic analysis

All sequences were aligned using SSE v.1.1 to available reference NPHV sequences from Genbank JQ434001-JQ434008 and CHV JF744991. NS3 helicase, NS5 and 5’UTR sequences from all NHV isolates were aligned with reference HCV sequences from all genotypes. Neighbour-joining trees for each genome region were constructed from Jukes-Cantor corrected pairwise distances calculated by using the program MEGA version 5 and datasets were bootstrap re-sampled 500× to indicate robustness of branching. The hepacivirus genotype 1a sequence, M62321, was used to root the tree (not shown). Scale bars indicate nucleotide substitutions per site.

4.2.6 Clinical history and liver enzyme data

Clinical histories for all actively viraemic horses were traced through participation with the Royal (Dick) School of Veterinary Sciences and animal owners. Diagnostic technicians at the Easter Bush Pathology department carried out liver enzyme analysis for all UK based samples. GGT, GLDH and bile acids were tested in serum of NPHV viraemic animals while ALT and AST were excluded as previously described in section 1.6.2.

4.3 Results

4.3.1 Screening for NPHV and homologs in range of mammalian species

To investigate the frequency of NPHV infection in dogs, 46 bronchoalveolar lavage (BAL) and/or saliva samples collected from dogs over a 6-month period in the Edinburgh area were screened by published PCR-based screening methods (Kapoor et al., 2011, Burbelo et al., 2012b) using primers targeting the 5’-UTR and NS3 regions. An RNA transcript from the NS3 region (Burbelo et al., 2012b) verified the sensitivity of the NS3-based assay to 5 RNA copies per amplification reaction (Table 4.1). All samples were negative in both genome regions (Table 4.2). Ninety-nine
plasma samples from dogs that had a variety of clinical conditions and had been referred for virology screening, along with 15 autopsy lung, liver, and spleen samples from dogs, were additionally screened; results were uniformly negative in both regions.

Since publication of the NS3- and 5′-UTR–based PCRs (Kapoor et al., 2011), comparative sequence data from several NPHV-infected horses became available (Burbelo et al., 2012b). These data revealed sequence variability in the primer binding regions of both primer sets. We therefore redesigned the screening primers in both regions (Table 4.2) to accommodate this variability. In the 5′-UTR region, it was additionally possible to ensure that primers matched homologous regions of HCV genotypes 1–7. The new nested NS3 primers showed similar sensitivity for the NS3 transcript (Table 4.1). The new NS3 primers were used to repeat screen all canine respiratory, plasma, and necropsy samples, and did not detect any active NPHV infections (Table 4.3).

To investigate the possible infection of non-canine mammalian species, we screened available plasma/serum, respiratory, and liver samples from horses (n = 175), donkeys (n = 16), domestic cats (n = 56), pigs (n = 40), and rodents (n = 61; 45 Apodemus sylvaticus, 8 Mus musculus, 5 Myodes glareolus, 3 Microtus agrestis) by using both sets of conserved primers (Table 4.3). From this extended survey, 3 plasma samples from horses were positive on initial screening and confirmed positive on re-extraction and re-amplification in 5′-UTR and NS3 regions. PCR of samples of all types from all other studied mammalian species showed negative results.
TABLE 4.3: Non-primate hepacivirus sequences detected by PCR on available mammalian samples from the United Kingdom*. A range of mammalian species were tested for the presence of CHV and NPHV RNA, and a range of sample types where available, in order to access the potential tissue tropism of the virus, where detected. Published primers sets were used in addition to study specific primers targeting the 5′UTR and conserved NS3 helicase of CHV and NPHV.

<table>
<thead>
<tr>
<th>Animal/sample type</th>
<th>No.</th>
<th>NS3</th>
<th>5′-UTR</th>
<th>Published NS3</th>
<th>New primers, 5′-UTR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dog</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Respiratory</td>
<td>53</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Plasma</td>
<td>99</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Lung</td>
<td>15</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Liver</td>
<td>15</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Spleen</td>
<td>15</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Horse</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Respiratory</td>
<td>40</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Plasma</td>
<td>328</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Donkey</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Plasma</td>
<td>100</td>
<td>NA</td>
<td>NA</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Cat</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Plasma</td>
<td>56</td>
<td>NA</td>
<td>NA</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Pig</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Serum</td>
<td>40</td>
<td>NA</td>
<td>NA</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Rodent</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Liver</td>
<td>61</td>
<td>NA</td>
<td>NA</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

* NS, nonstructural protein; UTR, untranslated region; NA, not applicable.

4.3.2 Phylogenetic analysis of NPHV variants with known members of the Hepacivirus genus

To confirm the presence of NPHV sequences in the 3 screen-positive horses, we further amplified each sample using conserved primers in the NS5B region and compared amplified sequences from each region with homologous regions of previously identified positive horses (Figure 4.1). Although this method does not represent a comprehensive genetic analysis, these sequence comparisons demonstrated that each of the positive horses was infected with NPHV variants distinct from the transcript-positive control and from each of the 8 previously identified infected horses in the USA and 4 infected horses in Germany. All 3 variants showed similar branching orders in each genome region (Figure 4.1),
consistent with the observed lack of recombination in previous analyses (Burbelo et al., 2012b).

\[ \text{FIGURE 4.1: Phylogenetic analysis of A) 5' untranslated region, B) nonstructural protein 3, and C) nonstructural protein 5B regions of nonprimate hepatitis virus sequence amplified from screen-positive study animals.} \]

Subsequent availability of novel hepacivirus isolates from rodents, bats and Old World monkeys allowed for an extended phylogenetic analysis of the Hepacivirus genus. Available sequences for RHV, BHV, GHV and NPHV were analysed with representative sequences from each HCV genotype, in the NS3 \((n=80)\) and NS5B \((n=39)\) regions and between nucleotides 4395-4576 and 8145-8300 respectively to maximise the dataset available for analysis. All NPHV isolates formed a distinct clade in both NS3 and NS5B with significant bootstrap support (Figure 4.2.). Variants isolated in this UK study were genetically distinct from all those previously identified in the US in addition to more recent isolates reported from Germany.

\[ \text{16 Sequence fragments in (A)5'UTR, (B) NS3 and (C) NS5B spanned nucleotide regions 125-355, 4395-4576 and 8145-8300 respectively, and were chosen to enable incorporation of all available NPHV isolates in analysis. Numbering of the genome is based on NPHV isolate Genbank no. JQ434003} \]
\[ \text{17 Sequence fragments analysed in Figure 4.2 (A) NS3 and (B) NS5B were shorter than the respective genomic regions enable incorporation of isolates from RHV, BHV, and GHV. Nucleotide regions 4395-4576 and 8145-8300 were applied in the NS3 and NS5B analysis respectively based on numbering of NPHV Genbank sequence JQ434003.} \]

159
(KC411810-KC411813). UK isolate JX948118 appears more closely related to the German isolates (marked green) in the NS3 region but a lack of available NS5B sequence data for the German samples does not allow this to be analysed further. All other UK isolates remain interspersed with those from the US study in both NS3 and NS5B regions.

All other isolates from rodents, bats and Old World monkeys fell into species specific bootstrap supported clades, with the exception of a single rodent isolate in the NS3 that grouped within the NPHV clade but lacked available comparable sequence data for the NS5B region. Despite the high genetic diversity all RHV and BHV isolates previously identified fell into distinct clades in both regions of the genome (Kapoor et al., 2013b, Drexler et al., 2013a, Quan et al., 2013). BHV isolates formed three clades in the NS5B region (Figure 4.2B), one most closely related to GBV-B and the remaining two formed a basal position relative to the NPHV, HCV and GHV clades, supporting findings also reported in a previous study (Drexler et al., 2013a). Comparable with previously published data isolates from rodents displayed high genetic diversity, the presence of at least five tentative viral species, and intraspecies genetic diversity similar to that observed between HCV subtypes, (Kapoor et al., 2013b) (Figure 4.2A). The basal positions of BHV and RHV suggested that these mammals might be ancient reservoirs for genetic diversity in the Hepacivirus genus.
NPHV detection and characterisation in domestic horses in the United Kingdom
The 3 infected horses originated in Scotland and comprised 2 geldings and 1 mare, between 12–20 years of age (Table 4.4). Clinical records for each horse from the time of sample collection failed to identify evidence of hepatitis or systemic disease. Samples were taken during routine veterinary screening (horse 1 and 3) or to test for potential environmental, bacterial or viral causes responsible for inflammatory airway disease (horse 2) (Table 4.4). Liver function tests carried out provided no evidence for hepatic inflammation: γ-glutamyl transferase (GGT) and glutamate dehydrogenase (GLDH) were within reference range, except a mildly elevated GGT level in horse 2 (Table 4.5). Hepatic insufficiency was also ruled out as bile acid levels were all within the reference range. Viral loads, measured by using real-time PCR against the NS3 transcript standard, ranged from 7 × 10⁴ to 5 × 10⁷ RNA copies/mL among the 3 horses (Table 4.5).

### TABLE 4.4: Clinical features of domestic horses infected by non-primate hepaciviruses in the United Kingdom.

All 3 NPHV viraemic horses were treated and referred with non-hepatic symptoms. Repeat samples were collected from horse 3 over a 4 month period and assessed for the progression of NPHV infection.

<table>
<thead>
<tr>
<th>Horse</th>
<th>Sample</th>
<th>Collection date</th>
<th>Area</th>
<th>Age, yr./sex</th>
<th>Presenting signs</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>EF_317/98</td>
<td>1998</td>
<td>Caithness</td>
<td>8/F</td>
<td>Lameness</td>
</tr>
<tr>
<td>2</td>
<td>EF_330/97</td>
<td>1997</td>
<td>Perthshire</td>
<td>12/M</td>
<td>Inflammatory airway disease</td>
</tr>
<tr>
<td>3</td>
<td>EF_369/11</td>
<td>2011 Dec 2</td>
<td>Lothian</td>
<td>20/M</td>
<td>Lameness; no lung disease</td>
</tr>
<tr>
<td></td>
<td>EF_374/12</td>
<td>2012 Mar 1</td>
<td></td>
<td></td>
<td>Lameness; no lung disease</td>
</tr>
<tr>
<td></td>
<td>EF_524/12</td>
<td>2012 Mar 23</td>
<td></td>
<td></td>
<td>Lameness; no lung disease</td>
</tr>
<tr>
<td></td>
<td>EF_725/13</td>
<td>2012 Dec 3</td>
<td></td>
<td></td>
<td>No symptoms</td>
</tr>
<tr>
<td></td>
<td>EF_726/14</td>
<td>2013 April 20</td>
<td></td>
<td></td>
<td>Lethargy</td>
</tr>
</tbody>
</table>
To investigate the clinical characteristics of NPHV infection, we further examined one of the NPHV positive horses (horse 3) and collected samples at 4, 5, 12 and 14 months after the original sample collection. The horse remained clinically unremarkable and had no specific symptoms indicative of systemic disease or hepatic illness. The horse regularly competed in equestrian events and had travelled worldwide extensively during the 10 years preceding the detection of infection with NPHV. Following close consultation with the owner no specific risk factors, such as operations, exposure to unsterilized needles, or history of systemic illness could be identified. During the initial 5-month follow-up period, the horse remained viraemic, but samples showed lower viral loads ($7 \times 10^4$ to $2 \times 10^5$ RNA copies/mL) than found in the initial sample ($5 \times 10^7$ copies/mL) (Table 4.4). At the time of final sampling viral RNA was undetectable, and while the horse exhibited signs of lethargy and fatigue there were no biological indicators of an association with NPHV or liver disease.

Liver indices were within the reference range for two of the three viraemic horses, with the exception of elevated GGT levels in horse 2, although liver enzymes and bile acids were frequently at the upper end of the reference range for all three horses. In the absence of elevated GLDH levels in combination with one or more of the other enzymes assayed it was not possible to infer that any hepatopathy or hepatic insufficiencies were present (as outlined in the criteria Section 1.6.1). The absence of detectable levels of RNA in horse 3 after 6 months is suggestive of acute resolving hepatitis infection. UK veterinary rules precluded us from taking a liver biopsy sample from horse 3 to further investigate the tissue tropism of the virus and site of viral replication. Nasal and mouth swab samples collected at month 5 in accordance with the detection of CHV in respiratory fluid and tested by PCR with both sets of primers, were NPHV negative. Peripheral blood mononuclear cells (PBMCs), a potential secondary site of HCV replication, were NPHV negative by PCR. Plasma from 6 horses stabled with horse 3 were uniformly NPHV PCR negative in 5’-UTR
and NS3. Serum taken from a single dog on the farm at month 14 also tested PCR negative for NPHV RNA.

**TABLE 4.5: Laboratory indices for domestic horses infected with nonprimate hepacviruses in the United Kingdom.** Liver enzyme analysis of actively viraemic horses was examined to determine the presence of underlying hepatic insufficiencies that might indicate viral replication in and damage to the liver. Elevated levels are marked with bold font.

<table>
<thead>
<tr>
<th>Horse</th>
<th>Sample</th>
<th>GGT&lt;sup&gt;1&lt;/sup&gt; (&lt;40 U/mL)</th>
<th>GLDH&lt;sup&gt;2&lt;/sup&gt; (&lt;10 U/mL)</th>
<th>Bile acids (&lt;10 U/mL)</th>
<th>Viral load, copies/mL</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>EF_317/98</td>
<td>15</td>
<td>2</td>
<td>1</td>
<td>1.3 × 10&lt;sup&gt;5&lt;/sup&gt;</td>
</tr>
<tr>
<td>2</td>
<td>EF_330/97</td>
<td>59</td>
<td>2</td>
<td>3</td>
<td>4.4 × 10&lt;sup&gt;5&lt;/sup&gt;</td>
</tr>
<tr>
<td>3</td>
<td>EF_369/11</td>
<td>15</td>
<td>2</td>
<td>6.4</td>
<td>4.8 × 10&lt;sup&gt;5&lt;/sup&gt;</td>
</tr>
<tr>
<td></td>
<td>EF_374/12</td>
<td>36</td>
<td>1</td>
<td>7.4</td>
<td>2.1 × 10&lt;sup&gt;5&lt;/sup&gt;</td>
</tr>
<tr>
<td></td>
<td>EF_523/12</td>
<td>24</td>
<td>4</td>
<td>6.3</td>
<td>7.1 × 10&lt;sup&gt;4&lt;/sup&gt;</td>
</tr>
<tr>
<td></td>
<td>EF_725/12</td>
<td>27</td>
<td>6</td>
<td>5.1</td>
<td>ND&lt;sup&gt;3&lt;/sup&gt;</td>
</tr>
<tr>
<td></td>
<td>EF_726/13</td>
<td>25</td>
<td>5</td>
<td>5</td>
<td>ND&lt;sup&gt;3&lt;/sup&gt;</td>
</tr>
</tbody>
</table>

<sup>1</sup>GGT, γ-glutamyl transferase.
<sup>2</sup>GLDH, glutamate dehydrogenase.
<sup>3</sup>ND, Non-detectable

### 4.4 Discussion

#### 4.4.1 Detection of Non-primate hepacivirus in UK Horses

Although the study was initially designed as an investigation of the frequency of CHV infection in dogs, initial findings of uniformly negative results from a large number of respiratory, plasma, and autopsy samples prompted us to widen our sampling to other mammalian species. Consistent with a then-recent report (Burbelo et al., 2012b), we detected the presence of viruses similar to CHV and now termed NPHV in ≈1% of horse plasma samples but absent from all other sample species tested (cats, pigs, or mice), irrespective of sample type (Table 4.3). The detection frequency of ≈1% (3/328) observed in this UK study was lower (significantly so by the Fisher exact test; p = 0.0007) than the previous US study which reported a viraemia frequency of 7.8% (8/103) among horses in New York state (Burbelo et al., 2012b).
The absence of detectable NPHV infection in cats, dogs, pigs and mice and the apparent restriction of NPHV infection to horses is consistent with serology-based screening carried out in the US (Burbelo et al., 2012b) that showed 35% of samples from horses contained antibodies against a recombinant NPHV NS3 peptide (of which ≈25% were additionally viraemic) but an absence of NS3 antibodies in all other species. With the exception of a single intermediately seroreactive sample from a cow all other samples from dogs, deer, rabbits and cows were seronegative in the US study. However, the single seroreactive cow was confirmed as NPHV RNA negative by PCR. These findings potentially suggested a rare infection in another ruminant and mammalian species or assay non-specificity that requires further investigation. Although these initial surveys provide preliminary evidence that horses could be the natural host of NPHV, its previous detection in dogs with respiratory disease (Kapoor et al., 2011) provides evidence for its potential spread to humans. In this respect, NPHV specificity would differ from the narrow specificity observed in other hepacviruses; e.g., HCV can infect humans and chimpanzees (although it does not naturally circulate in the latter species) and cannot infect Old World monkey species, such as macaques (Bukh et al., 2001a). The host range of GBV-B appears similarly restricted to New World monkeys (Bukh et al., 2001b).

In 2013 the unprecedented detection of novel and highly divergent hepacviruses in rodents, bats and New World monkeys has dramatically impacted hypotheses on the origins of human HCV (Kapoor et al., 2013b, Drexler et al., 2013a). In a singe large global study of sera and organs from 4,470 rodents, RHV was detected in 27 (1.8%) of 1,465 European bank voles (Myodes glareolus) and 10 (1.9%) of 518 South African four-striped mice (Rhabdomys pumilio) (Drexler et al., 2013a). An anti-HCV immunoblot carried out as part of the same study detected reactivity in 13 bat samples (7.2%, 9 R. aegyptiacus and 4 E. helvum) out of 2,939 tested. However, hepacivirus RNA was not detected in any seroreactive bats comparable to the seroreactive but NPHV RNA negative cow previously discussed (Kapoor et al., 2011).
2013b). Whether the seroreactivity observed to NPHV and HCV in cows and rodents respectively was the consequence of assay non-specificity or an indicator of the presence of divergent hepaciviruses amongst diverse mammalian species remains to be determined. The concomitant detection of divergent hepaciviruses in 0.6% of bats sampled by a second group appears to support the aforementioned study by Drexler et al. indicating that highly divergent hepaciviruses are present in a range of mammalian species (Quan et al., 2013, Drexler et al., 2013a). The detection of divergent and novel hepaciviruses in a range of mammalian species across the globe provides valuable information and insight into the origins of human HCV. Further screening of an extended range of mammalian species and sub-species around the world, in particular in regions of Africa associated with endemic HCV infection may help to elucidate the ultimate origins.

4.4.2 Genetic relationship between NPHV and other members of the Hepacivirus genus

Based on phylogenetic analysis all UK NPHV isolates are distinct from those previously reported in the US and Germany (Figure 4.1, 4.2). Nucleotide divergence of study isolates from those of the US and Germany ranges between 7.6-18.2% and 8-11.9% respectively, while that observed between isolates ranges between 3.2-7.2%. The divergence observed between all isolates of NPHV is greater in the S and NS regions of the genome than that observed between HCV subtypes and between variants of HPgV suggesting the presence of circulating subtypes of NPHV in the horse population. Analysis of NPHV isolates with novel RHV and BHV sequences implies that bats and rodents may harbour ancient sources of hepacivirus genetic diversity, and further screening of these species in the UK and endemic regions of HCV infection in Africa will establish more clearly the genetic relationship between novel members of the Hepacivirus genus and HCV. Despite the genetic diversity of RHV and BHV isolates, distinct species-specific clades are identifiable in NS3 and NS5B regions analysed (Figure 4.2) and suggest the occurrence of successful
transmission between species and subsequent adaptation and evolution of the virus within host species. Increased screening and availability of sequences across species will direct the development of a conclusive hypothesis for the evolution and dissemination of hepaciviruses worldwide and determine whether bats or rodents are the ancient natural reservoir of hepaciviruses.

4.4.3 Clinical presentations of NPHV infections

Neither the clinical signs nor the liver function tests of the 3 NPHV-infected horses provided a clear indication of the organism’s association with hepatic or other systemic disease (Tables 4.4, 4.5). The liver is the primary site of HCV and GBV-B replication in naturally infected individuals and experimentally infected New World monkeys respectively. Viral genomes can also be detected at lower level in peripheral mononuclear cells in some hosts. Diagnosis of acute HCV infection is infrequently made as more than 70% of patients are asymptomatic with acute infection. The remainder of patients experience jaundice, and develop gastrointestinal problems. When symptoms do occur they typically present at 6 to 8 weeks post exposure and last between 2 to 12 weeks. Examination of the clinical histories of all horses in this study did not indicate any systemic disease, or possible symptoms attributed to hepatic disease in horses. Symptoms frequently observed include weight loss, hepatic encephalopathy, icterus, hypoglycaemia, haemoglobinuria, while less frequently observed signs are photosensitivity, bleeding abnormalities, oedema and ascites (Durham et al., 2003a, McGorum et al., 1999, Cornick et al.).

GGT and GLDH are sensitive markers of liver inflammatory processes in the horse. Previous studies have applied the use of GGT, GLDH and bile acids as diagnostic tools in examining hepatic disease in horses (Divers, 1998, Divers, 2005, Durham et al., 2003a). A single study of 50 horses with hepatic disease documented significantly elevated levels of GGT, GLDH and bile acids and further determined
that levels were significantly higher in horses which were euthanized or died as a consequence of liver disease (McGorum et al., 1999). Studies presented within this chapter found all enzyme levels were within the reference range for 2/3 of the viraemic horses, horse 2 was observed to have elevated GGT 59 (Table 4.5). Reference levels of bile acids similarly demonstrated adequate liver function. Although the sample size was small, these relatively normal liver indices contrast with the frequent GGT and ALT elevations associated with chronic HCV infection (Hwang et al., 2001, Kiso et al., Silva et al., 2004, Oberti et al., 1997, Bonacini et al., 1997) and found in New World monkeys experimentally infected with GBV-B (Beames et al., 2001, Iwasaki et al., 2011, Jacob et al., 2004, Lanford et al., 2003, Martin et al., 2003). In most cases of HCV infection the initial indicators of infection is from laboratory abnormalities. Elevations in alanine aminotransferase (ALT) are frequently observed 40 to 50 days post infection or between 6 and 112 days in some cases of post transfusion HCV.

Although in the current study, UK veterinary regulations did not permit liver biopsies to be performed on horses without evidence of liver disease, the current findings do not rule out a lower grade infection or potential replication in the liver without the associated inflammatory response to HCV that is primarily responsible for liver damage (Klenerman and Thimme, 2012, Guidotti and Chisari, 2006).

**4.4.4 Evidence for establishment of persistent infection**

High viral loads were detected in all three horses in this study ($1.3 \times 10^5$-$4.8 \times 10^7$ RNA copies/ml) and exceeded or fell within the range observed for HCV and GBV-B respectively. All horses were clinically unremarkable despite the high viral loads, comparable with recent data from bats indicating that all animals appeared healthy despite qPCR indicating $10^3$ to $10^8$ RNA copies/mL (Quan et al., 2013). Serum HCV and GBV-B generally peaks at 6 to 10 weeks regardless of the progression to chronic
infection or resolved, levels of virus are present in the blood, ranging between 1-10 million and 10-100 million genome equivalents/ml, respectively.

The detection of NPHV RNA sequences in samples obtained 5 months apart from horse 3 provides evidence for an ability of NPHV to establish persistent infections. Acute HCV is typically taken to refer to the initial 6 months of infection. The course of the disease is typically defined as occurring over 3 stages, pre-ramp up phase, ramp up phase and the plateau phase. The ramp up phase sees the exponential replication of HCV, and a doubling of viral load approximately every 10 hours.

The decline and subsequent clearance of detectable levels of viral RNA after 6 months in horse 3 and throughout the remaining sampling period, indicates infection occurred within a year of the first sampling. This finding is suggestive of an acute spontaneously resolved infection. Based on liver enzyme levels reported previously there is no indication of chronic infection or underlying systemic disease (Lyons et al., 2012a). The subsequent clearance of viraemia is suggestive of an acute self-limiting infection, comparable with some cases of persistent HCV viraemia associated with acute self-limiting post-transfusion hepatitis (Barrera et al., 1995). Although it is not possible to determine the exact time of initial infection it is likely that the first sampling in December 2011 was taken after recent exposure to NPHV, at or shortly after the peak in viral load and overlapping seroconversion. In contrast with the estimated 70-90% progression to chronic HCV infection (Mondelli et al., 2005, Heller and Rehermann, 2005), NPHV has not yet been documented to result in chronic infection or progression to liver cirrhosis and cancer.

Although longer-term sampling of NPHV viraemic horses is required to confirm whether the virus establishes a persistent infection, the detection of viral RNA in samples 5 months part is consistent with the high viraemia frequencies among seropositive horses in a previous study (8/37) (Burbelo et al., 2012b). This proportion would probably not be observed on random sampling if infections had
rapidly resolved, and implies that NPHV can establish persistent infections. Whether these infections proceed to a chronic state as observed in approximately 70% of HCV infections which in the absence of treatment leads to decades- or life-long viraemia and active liver disease in >50% of individuals, remains to be determined. In contrast GBV-B although clearly hepatotropic, does not establish persistent infections in tamarins or owl monkeys (Muerhoff et al., 1995, Bukh et al., 2001b, Beames et al., 2000). However, more recent studies have demonstrated long-term persistence among experimentally infected marmosets (Ishii et al., 2007, Iwasaki et al., 2011).

4.4.5 Conclusions

Data presented here provides preliminary clinical and virological insight into the course of NPHV infection. Based on the range of mammalian species tested we can infer that horses appear to be the natural hosts for NPHV. Furthermore the failure to detect CHV in any canine liver, serum or BAL samples screened indicates that the initial detection and characterization of CHV in a respiratory outbreak in the US may have been a rare and possibly self-limiting transmission event between species. Wider sample screening in different countries and amongst dogs in contact with horses will yield further insight into this possibility.

While no correlation can be drawn from this data as to the tissue tropism or disease association of the virus in horses, it provides initial evidence for the development of persistent infections and tentative results suggesting that there is no direct association between NPHV and hepatopathy or the development of hepatitis in horses. This is based on the observation that the liver enzymes assayed were not sufficiently elevated alone or in combination to indicate any underlying hepatopathy as outlined in criteria in section 1.6.1. Future studies, perhaps refocused on NPHV screening of horses with idiopathic liver disease that have undergone biopsy sampling and have been clinically characterized, are needed to investigate further the potential for hepatotropic NPHV and manifest its clinical effects. In the longer term, and
acknowledging that the horse is not the ideal experimental animal, inoculation of horses with NPHV and subsequent monitoring for viraemia development, liver function abnormalities, and B- and T-cell immune responses would provide further insights into the nature of NPHV infections and associated immune response and similarity of these developments to current observations for HCV and GBV-B. The development of novel serological assays to determine previous exposure across diverse mammalian species will provide insight into the possibility of NPHV transmission between species. Furthermore the screening of bats and rodents in the UK for novel hepaciviruses may provide valuable information relating to transmission routes, dissemination, tissue tropism and evolutionary history of NPHV, and as a consequence fill in valuable pieces of the puzzle with respect to our understanding of the ultimate origins of HCV in humans.
Chapter 5
Epidemiology, Clinical Presentations and structural analysis of 5’UTR of EPgV

5.1 Introduction

5.1.1 Host range and genetic diversity of members of the Pegivirus genus

As previously discussed in section 1.5.1 pegiviruses infect a diverse range of mammals including humans [HPgV] (Simons et al., 1995a, Simons et al., 1995b, Yoshida et al., 1995), bats [BPgV](Epstein et al., 2010), new world monkeys and chimpanzees [SPgV/SPgVcp] (Adams et al., 1998, Birkenmeyer et al., 1998, Bukh and Apgar, 1997, Epstein et al., 2010, Kapoor et al., 2013a, Leary et al., 1996a, Linnen et al., 1996, Muerhoff et al., 1995, Quan et al., 2013, Simons et al., 1995b). Recently divergent pegiviruses have been isolated in two separate studies in horses in the United States (Kapoor et al., 2013a, Chandriani et al., 2013). These two distinct equine pegiviruses have been termed equine pegivirus (EPgV) (Kapoor et al., 2013a) and Theiler’s disease associated virus (TDAV) (Chandriani et al., 2013). Theiler’s disease is an acute serum hepatitis in horses associated with the administration of blood products, for example botulinum antitoxins (Hjerpe, 1964, Pearson, 1999). The Chandriani et al. study concluded that TDAV may be the aetiological agent responsible for theiler’s disease and is potentially transmitted between horses in contaminated blood derived products (Chandriani et al., 2013).

Recent global studies of wild rodent and bat populations has also detected rodent pegivirus infections [RPgV] and further BPgV isolates, revealing a far greater viral diversity in members of the Pegivirus genus (Kapoor et al., 2011, Kapoor et al., 2013a, Burbelo et al., 2012b, Quan et al., 2013, Drexler et al., 2013a, Lauck et al., 2013). Currently the lack of available sequence data covering comparable regions of BPgV and RPgV genomes currently places limitations on the regions and/or numbers
Structural analysis, epidemiology and clinical characteristics of EPGV infection

5.1 Epidemiology, Clinical Presentations and structural analysis of 5’UTR of sequences available for phylogenetic analysis. Where sequence data is available species specific clustering of all pegivirus isolates is observed by phylogenetic analysis of NS3 and NS5B regions despite the observed genetic diversity of recent rodent and bat-derived sequences (Quan et al., 2013, Kapoor et al., 2013a).

5.1.2 Frequency and distribution of pegivirus infection in humans and NHPs

Infections associated with HPgV are globally distributed with levels of active viraemia estimated at ≈5%, with antibodies detectable in 5-13% of blood donors in developed countries (Blair et al., 1998, Gutierrez et al., 1997, Pilot-Matias et al., 1996a, Tacke et al., 1997a). Rates of viraemia in developing countries amongst blood donors can be as high as 20% in some regions (Polgreen et al., 2003, Mohr and Stapleton, 2009, Pavlova et al., 1999). Significant increases in HPgV infections have been reported amongst individuals with sexually transmitted or blood-borne diseases (Scallan et al., 1998b), a single study reported active viraemia in 39.6% of cases (Williams et al., 2004). Infection frequencies amongst bats, rodents and horses are estimated at 3.9-5% (Quan et al., 2013, Kapoor et al., 2013a, Kapoor et al., 2013b). The primary study on the prevalence of EPGV infection observed active viraemia in 25% (3/12) of horses with elevated liver enzymes and 6.4% (4/62) among healthy animals (Kapoor et al., 2013a). Based on the sample size bias no correlation between EPGV infection and liver enzyme levels could be drawn (p=0.08). Eight out of 22 (36.4%) horses in a separate study were found to be infected with novel equine pegivirus TDAV following receipt of a botulinum antitoxin (Chandriani et al., 2013), suggesting potential transmission route for the virus.

5.1.3 Clinical indications of pegivirus infection

Pegiviruses originally termed ‘GB’ viruses were thought to be associated with hepatic disease. The reclassification of BPGV (GBV-D), SPgV (GBV-A), EPGV and RPGV as
members of the *Pegivirus* genus is a reflection of a lack of evidence suggesting an association with hepatitis (Stapleton *et al.*, 2011). Until recently, the only other virus outside HCV in the *Hepacivirus* genus was GBV-B, detected in a laboratory tamarin (Simons *et al.*, 1995a, Muerhoff *et al.*, 1995) in which it caused hepatitis and could be experimentally transmitted to New World primate species (Bukh *et al.*, 2001).

The clinical significance of infection with BPgV (GBV-D), SPgV<sub>cpz</sub> (GBV-C<sub>cpz</sub>), SPgV (GBV-A), EPgV and RPgV has not been determined. Current findings predominately suggest no association between hepatitis and HPgV infection (Theodore and Lemon, 1997, Feucht *et al.*, 1997) and research has shown that co-infection with HPgV inhibits HIV replication (Timmons *et al.*, 2013, Polgreen *et al.*, 2003, Stapleton, 2003, Williams *et al.*, 2004, Xiang *et al.*, 2001). SPgV and HPgV viruses are present in low to non-detectable levels in the liver of infected hosts, and virus can be more easily detected in lymphocytes suggesting both viruses are not hepatotrophic but lymphotrophic (Kobayashi *et al.*, 1999, Laskus *et al.*, 1997, Pessoa *et al.*, 1998, Radkowski *et al.*, 1999, Radkowski *et al.*, 2000, Simons *et al.*, Tucker *et al.*, 2000).

BPgV has been detected from clinically healthy bats at concentrations between $10^3$-$10^8$ RNA copies per ml of plasma, although levels in the liver and other cells have not been analysed (Quan *et al.*, 2013, Epstein *et al.*, 2010). Viral loads of between $10^{4.5}$ and $10^{6.5}$ were detectable in the plasma of horses infected with EPgV, and RNA was further detected in the liver, lymph and PBMCs of two chronically infected horses but there was no significant difference in RNA levels between the tissues (Kapoor *et al.*, 2013a). TDAV infection contrasts with other members of the *Pegivirus* genus as findings from three separate outbreaks of acute serum hepatitis suggest an association with liver disease in horses (Chandriani *et al.*, 2013). Further study is necessary to determine the exact tissue tropism associated with EPgV infection in horses.
5.1.4 Progression of pegivirus infection in the horses

Studies by Chandriani and Kapoor in the US in the last year provide evidence for the establishment of acute and chronic pegivirus infections. EPgV was detected in two horses from a healthy herd over a period of at least 3.5 years suggesting chronic infection and RNA was detectable in liver, lymph, PBMCs and plasma (Kapoor et al., 2013a). Acute resolving infections were also observed in four animals in the same the study. In contrast Theiler’s disease associated virus (TDAV), a second genetically distinct equine pegivirus, has been proposed as the etiological agent responsible for acute hepatitis in horses following the administration of blood products (Chandriani et al., 2013). TDAV was detected in 8/22 horses in the study and subsequent inoculation of 4 healthy horses with TDAV-positive plasma resulted in several weeks of viraemia preceding liver disease, with one horse presenting with elevated liver enzymes.

5.1.5 Pegivirus 5’UTR: Secondary structure and type 4 internal ribosomal entry site (IRES)

Like other members of the Flaviviridae family, HPgV, SPgVcpz, and SPgV contain a 5’UTR with an internal ribosomal entry site (IRES) that directs translation of the polyprotein from viral genomic RNA (Simons et al., 1996, Yoo et al., 1992). All pegiviruses are predicted to contain an IRES element, although their structures differ. Structural predictions for SPgV and HPgV conform with a type 4 IRES element (Kieft, 2008). Initial structural predictions of the 5’UTR of EPgV suggest a type 1 IRES structure similar to that present in picornaviruses (Kapoor et al., 2013a, Alexander et al., 1994, Belsham, 2009, Fernandez-Miragall et al., 2009). Studies on the 5’UTR secondary structure of BPgV and RPgV have not been carried out. Currently studies are limited by the availability of comparable sequence data to definitively determine IRES type of rodent, bat and equine pegiviruses.
In the results presented in this chapter we have sought to determine the frequency of active EPgV and TDAV infection among horses in the United Kingdom and to identify possible associations with liver and other diseases in horses through analysis of liver enzymes and serum viral load. Polymerase chain reaction (PCR)-based assays for direct viral RNA detection were used to screen a wide range of other mammalian species to re-investigate the host range and species-specificity associated with both pegiviruses. Phylogenetic analysis of the evolutionary conserved NS3 helicase and NS5B RNA polymerase was used to establish the genetic relationship between EPgV isolates of this study and those previously reported. Sequences of the 5’UTR of EPgV isolates in this study in addition to those already available allowed for secondary structural predictions and examination of IRES structure. Sequences reported in this study were assigned GenBank numbers KF923465-KF923496.

5.2 Materials and Methods

5.2.1 Sample selection

Serum or plasma samples (n=835) were collected between 1995 and 2013 from 163 NHPs (Cameroon), 328 horses (Scotland, England and France), 100 donkeys (England), 113 dogs (Scotland) and 131 cats (Scotland). Samples from NHPs were collected in Cameroon by Metabiota (formerly Global Viral Forecasting Initiative) staff in Cameroon and included 11 gorillas, 62 chimpanzees and 91 Old World Monkeys (OWM) as previously outlined (Lavoie et al., 2012; Lyons et al., 2012b; Sharp et al., 2010). Blood samples were collected via venepuncture from 73 apes comprising 11 gorillas (Gorilla gorilla) and 62 chimpanzees (Pan troglodytes troglodytes and Pan troglodytes elliotti), and from a variety of Old World Monkey species: Cercocebus agilis (n = 7), C. torquatus (n = 2), Cercopithecus cephus (n = 3), C. erythrotis (n = 4), C. l'hoesti preussi (n = 4), C. mona (n = 9), C. nictitans (n = 3), C. pogonias (n = 1), C. tantalus (n = 3), Erythrocebus patas (n = 3), Lophocebus
Structural analysis, epidemiology and clinical characteristics of EPgV infection
Epidemiology, Clinical Presentations and structural analysis of 5’UTR

albigena (n = 5), Mandrillus leucophaeus (n = 20), M. sphinx (n = 9), and Papio anubis (n = 20). If not previously recovered serum was separated from whole blood by centrifugation and stored long term at -80ºC. Samples were shipped to the United Kingdom from Cameroon in compliance with UK and Cameroon laws and the Convention on International Trade in Endangered Species of Wild Fauna and Flora (CITES) (Sharp et al., 2010).

Samples from horses, donkeys, dogs and cats were archived excess diagnostic samples collected at the Royal (Dick) School of Veterinary Studies, Edinburgh, Liphook Equine Hospital Laboratory, Hampshire and The Donkey Sanctuary, Sidmouth. Animals were of mixed breed, age and sexes. Donkeys, dogs and cats had a wide range of diseases that prompted diagnostic sampling. Horses were categorised into a hepatopathy group (n=111) and a control group (n=217). The hepatopathy group comprised horses that had been investigated for suspected liver disease and which had biochemical evidence of hepatopathy, as indicated by serum/plasma concentrations of GGT, GLDH and/or bile acids exceeding the laboratory reference upper level. Hepatopathy was attributed to a wide range of causes. The control group comprised horses for which there was no clinical suspicion of hepatopathy, including clinically healthy horses and horses that were sampled for investigation of a wide range of diseases excluding hepatopathy. Serum samples were collected between 1995 and 2013 and anonymised prior to testing.

5.2.2 NPHV, EPgV, and TDAV Screening

All samples were screened for NPHV by PCR with previously published primers and transcript control (Lyons et al. 2012). Horse and donkey samples were screened for EPgV and TDAV RNA. For detection of EPgV RNA, nested PCR primers were designed targeting the viral NS3 helicase based on published sequences and used in addition to published primer sets (Burbelo et al., 2012b). To validate the EPgV and
TDAV PCR, RNA transcripts were generated from a plasmid containing PCR amplified partial NS3 cDNA by using the Ambion T7 transcription kit (Promega Corp., Southampton, UK). Transcripts were purified with the RNeasy kit (QIAGEN), and concentrations were determined by using the NanoDrop 2000 (NanoDrop Products, Wilmington, DE, USA). RNA extractions were performed on 140 µl of plasma using the QIAmp viral extraction kit (QIAGEN) according to the manufacturer’s instructions and eluted in a final volume of 60 µl.

RNA was converted to cDNA using random hexamers with the Superscript III Reverse Transcription System (Life Technologies) and then used in nested PCR with previously published NPHV NS3 primers with newly designed EPgV NS3 and 5’UTR, and TDAV-NS3 primer sets (Appendix A) and amplified using 2 rounds of 35 cycles at 94°C for 18 s, 50°C for 21 s, and 72°C for 1.5 min; and 1 cycle of 72°C for 5 min, with 2 µL of first-round amplicon added to the second round. TDAV qPCR screening was carried out as previously published (Chandriani et al., 2013). The EPgV and TDAV NS3 transcript was tested by using both NS3 primer sets (Appendix A) and used as a control in screening, with sensitivity of 0.5–5 RNA copies in a reaction (Table 5.1). Positive second-round PCR amplicons were sequenced as previously published (Lyons et al., 2012) and analysed using SSE v1.1 (Simmonds, 2012a).

**TABLE 5.1: Transcript titration of NPHV and EPgV NS3 helicase controls tested with virus specific NS3 primers.** Serial dilutions of NPHV and EPgV NS3 helicase transcript controls were tested with relevant primers to determine the sensitivity of the PCR in detecting viral RNA isolated from serum samples. Sensitivity of both primers sets was determined to between 5 and 0.5 RNA copies per ml.

<table>
<thead>
<tr>
<th>Transcript RNA copies/mL</th>
<th>NPHV-NS3</th>
<th>EPgV-NS3</th>
</tr>
</thead>
<tbody>
<tr>
<td>5 x 10⁶</td>
<td>2/2</td>
<td>2/2</td>
</tr>
<tr>
<td>5 x 10⁵</td>
<td>2/2</td>
<td>2/2</td>
</tr>
<tr>
<td>5 x 10⁴</td>
<td>2/2</td>
<td>2/2</td>
</tr>
<tr>
<td>5,000</td>
<td>2/2</td>
<td>2/2</td>
</tr>
<tr>
<td>500</td>
<td>2/2</td>
<td>2/2</td>
</tr>
<tr>
<td>50</td>
<td>2/2</td>
<td>2/2</td>
</tr>
<tr>
<td>5</td>
<td>2/2</td>
<td>2/2</td>
</tr>
</tbody>
</table>
5.2.3 Viral load quantification

Real-time quantitative PCR (qPCR) was used to determine viral loads of positive samples and a standard calibration curve was generated from a dilution series of the NS3 transcript. Dilutions of EPgV and TDAV NS3 transcript were prepared from concentrations of $10^6$ to 1 copy/µL; 5 µL of transcript RNA was used to generate cDNA using random hexamers and reverse transcription. Five µL aliquots of cDNA were assayed in triplicate for the positive samples in the same way. To quantify viral loads of positive samples, EPgV NS3IS and EPgV NS3IAS primers were used with 4 µL of cDNA in the SensiFAST SYBR Hi-ROX Kit (BioLine, London, UK) per manufacturer’s instructions with the exception that the annealing temperature was reduced to 50°C and the extension time extended to 15 s. Samples were analysed in triplicate and fluorescence measured by using the Rotor-Gene Q system (QIAGEN). Viral loads were read from the standard curve generated and converted to RNA copies/mL based on the sample volume used in extraction and elution of the RNA.

5.2.4 Clinical history and liver enzyme data

Clinical histories for all actively viraemic horses were traced through participation with the Royal (Dick) School of Veterinary Sciences and animal owners. Diagnostic technicians at the Easter Bush Pathology department carried out liver enzyme analysis for all UK based samples. GGT, GLDH and bile acids were tested in serum or plasma of NPHV and EPgV viraemic animals (where sufficient sample was available), alanine transaminase and aspartate transaminase were considered less informative and excluded from liver enzymes tests as previously discussed (Chapter 2).
5.2.5 Determination of 5’UTR secondary structure
The secondary structure of the EPgV 5’UTR was determined by combined application of ALIFOLD, PFOLD and StructureDist SSE v1.1 software as detailed in Section 2.3.7. Where available complete fragments of 620 nucleotides were used from all available Genbank EPgV 5’UTR sequences in addition to all those obtained in this study\(^{18}\).

5.3 Results
5.3.1 Hepacivirus infections in non-human primates
We previously reported the detection of NPHV infection in three horses in a sample of 142 drawn from horses across Scotland. Samples from all other mammalian species screened (99 dogs, 158 horses and donkeys), 56 cats (*Felis catus*), 63 rodents (*Apodemus sylvaticus*, *Mus musculus*, *Myodes glareolus*, *Microtus agrestis*), and 40 pigs (*Sus scrofa*) were NPHV RNA-negative (Lyons et al., 2012).

NPHV primer sets previously published (Kapoor et al., 2011, Kapoor et al., 2013a, Lyons et al., 2012a, Chandriani et al., 2013) were applied to screen a larger study group of horses and other mammalian species (328 horses, 100 donkeys, 116 dogs, 132 cats, 362 human, 164 non-human primates, NHP) (Table 5.2). No NPHV infections additional to those previously reported (Lyons et al., 2012a) were detected in the larger equine sample set. All human samples and NHPs tested were also negative for NPHV RNA. Testing of NHPs with additional degenerate NS5B and 5’UTR primers sets targeting all genotypes of HCV detected no hepativirus RNA (Appendix A).

\(^{18}\) Genbank sequences used in the modelling of EPgV 5’UTR were Genbank IDs NC_020902.1 and KC410872 in addition to the available sequence C35. Complete study sequences used were EF_381/97, EF_383-385/97. Partial study sequences included were EF383, EF697, EF588, EF601, EF604, EF556, EF610 and EF593.
**TABLE 5.2: Detection of NPHV, EPgV and TDAV RNA in non-human primates.** A range of non-human primates were screened for the presence of hepatitis C and pegivirus infections applying previously published primer sets (Kapoor et al., 2013a, Lyons et al., 2012a, Kapoor et al., 2011) and additional study primer sets (Appendix A).

<table>
<thead>
<tr>
<th>Species</th>
<th>NPHV</th>
<th>EPgV</th>
<th>TDAV</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>PCR+</td>
<td>PCR+</td>
<td>PCR+</td>
</tr>
<tr>
<td>Horses:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hepatopathy</td>
<td>3/328</td>
<td>12/327</td>
<td>0</td>
</tr>
<tr>
<td>Non-hepatopathy</td>
<td>0/111</td>
<td>7/111</td>
<td>0</td>
</tr>
<tr>
<td>Donkeys</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0/100</td>
<td></td>
<td></td>
<td>0</td>
</tr>
<tr>
<td>Humans:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hepatopathy</td>
<td>0/362</td>
<td>0/96</td>
<td>ND</td>
</tr>
<tr>
<td>Non-hepatopathy</td>
<td>0/266</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-human primates</td>
<td>0/164</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dogs</td>
<td>0/113</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cats</td>
<td>0/131</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\(^1\text{ND; not done}\)

### 5.3.2 Pegivirus infections in horses

To determine frequencies of active EPgV and to investigate the relationship between viraemia and liver disease, all samples were screened by PCR using EPgV-specific primers targeting the NS3 region (Table 5.2) (Kapoor et al., 2013a). A transcript control based on the NS3 helicase was generated from a high titre sample and verified the sensitivity of the NS3 based PCR assay to 5 RNA copies/reaction (Table 5.1). Active infection was detected in 12/328 horses (Table 5.2). Samples were collected between 1997 and 2013 from range of collection locations. Further PCRs based on the published 5’UTR and NS5B of EPgV confirmed the presence of viraemia in all positive samples. All PCR products were sequenced and their genetic relatedness compared to published sequences of EPgV, TDAV (Chandriani et al., 2013, Kapoor et al., 2013a) and the recently isolated RPgV and BPgV (Figure 5.1). All EPgV isolates
were distinct from those isolated in the US and from the single available TDAV sequence isolated from an outbreak of acute serum hepatitis in horses (Figure 5.1).

5.3.3 Phylogenetic analysis of novel EPgV sequences in conserved NS3 and NS5B regions of the pegivirus genome

Samples testing RNA positive for EPgV by screening PCR in NS3 region using previously published primer sets (Kapoor et al., 2013a) were sequenced and further amplified in the NS5B RNA polymerase and 5’UTR regions to confirm positives. All available Genbank sequences for EPgV, RPgV and BPgV were included in analysis and divergent sequence sets were chosen from among available HPgV and SPgV sequences (Appendix C). Phylograms were rooted using NPHV sequence of GenBank accession number JQ434002.1. Shorter fragments of the NS3 helicase and NS5B RNA polymerase were used in the analysis to maximise the number of sequences available. Phylograms of the NS3 and NS5B region were constructed using nucleotide fragment 4603-4802, and 9251-9540 respectively based on numbering of EPgV Genbank sequence NC_090902. Phylogenetic analysis confirmed all study isolates as distinct from those previously reported for EPgV and TDAV in both the NS3 and NS5B regions (Figure 5.1, 5.2) (Chandriani et al., 2013, Kapoor et al., 2013a).

Distinct clustering of all variants into bootstrap supported species-specific clades is also evident in both NS3 and NS5B. The sequence diversity observed for BPgV is evident by the presence of three distinct lineages in the NS3 and NS5B region. An RPgV specific clade was predicted in the NS3 but the lack of availability of parallel sequences for the NS5B (Figure 5.2) limits the number of rodent derived sequence to just one in this region, which fell within a BPgV clade.

Sequence analysis of all study EPgV sequences in the NS3 and NS5B determined nucleotide divergence of 5.5% and 5.6%, and amino acid divergence of 0.5% and 0.6% between isolates, respectively. Comparative sequence analysis between EPgV
and pegiviruses of other species BPgV/RPgV/SPgV and GHV in the NS3 and NS5B region determined nucleotide divergence of between 40-53% and 39-47.5% and amino acid diversity between 46-59% and 33-51.7% respectively (Table 5.3).

Table 5.3: Nucleotide and amino acid (shaded) divergence in the NS3 and NS5B regions between EPgV isolates and known sequences from each of the species-specific clades; RPgV, BPgV, SPgV and HPgV. NS3 and NS5B regions analysed are comparative to those in Figure 5.1 and 5.2, and cover nucleotide fragments 4603-4802 and 9251-9540 respectively, numbering based on Genbank sequence EPgV NC020902.

<table>
<thead>
<tr>
<th>Virus</th>
<th>NS3</th>
<th>NS5B</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Amino Acid</td>
<td>Nucleotide</td>
</tr>
<tr>
<td>EPgV</td>
<td>53%</td>
<td>46.3%</td>
</tr>
<tr>
<td>BPgV</td>
<td>49.3%</td>
<td>46.9%</td>
</tr>
<tr>
<td>RPgV</td>
<td>49.4%</td>
<td>47.3%</td>
</tr>
<tr>
<td>HPgV</td>
<td>49.1%</td>
<td>46.2%</td>
</tr>
<tr>
<td>SPgV</td>
<td>48.4%</td>
<td>45.5%</td>
</tr>
<tr>
<td>TDAV</td>
<td>40.1%</td>
<td>41.3%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Virus</th>
<th>NS3</th>
<th>NS5B</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>50.1%</td>
<td>41%</td>
</tr>
<tr>
<td></td>
<td>49.8%</td>
<td>44.9%</td>
</tr>
<tr>
<td></td>
<td>51%</td>
<td>45.3%</td>
</tr>
<tr>
<td></td>
<td>48.1%</td>
<td>43.6%</td>
</tr>
<tr>
<td></td>
<td>49.7%</td>
<td>39.6%</td>
</tr>
<tr>
<td></td>
<td>44%</td>
<td>39.3%</td>
</tr>
<tr>
<td></td>
<td>45.1%</td>
<td>35.3%</td>
</tr>
<tr>
<td></td>
<td>44%</td>
<td>35.3%</td>
</tr>
<tr>
<td></td>
<td>46%</td>
<td>35.3%</td>
</tr>
<tr>
<td></td>
<td>39%</td>
<td>47.3%</td>
</tr>
<tr>
<td></td>
<td>38.7%</td>
<td>33.5%</td>
</tr>
<tr>
<td></td>
<td>41.9%</td>
<td>33.5%</td>
</tr>
</tbody>
</table>
Structural analysis, epidemiology and clinical characteristics of EPgV infection

Epidemiology, Clinical Presentations and structural analysis of 5’UTR
FIGURE 5.1: Phylogenetic analysis of EPgV study isolates in the NS3 helicase region with representative members of the Pegivirus genus. Maximum-Likelihood trees of nucleotide sequences were constructed using Tamura-Nei model of estimated distances calculated by using the program MEGA version 6.06 (datasets were bootstrap re-sampled 500× to indicate robustness of branching [values >70% shown on branches]). The best-fit model for the data was determined using Mega 6.06 best substitution model test. Trees were rooted using NPHV G1_073_GBX2. Scale bars indicate nucleotide substitutions per site. Study sequences are marked in bold. 19

19 Based on the numbering of EPgV NC020902 nucleotide fragment 4603-4802 was used.
Structural analysis, epidemiology and clinical characteristics of EPgV infection

Epidemiology, Clinical Presentations and structural analysis of 5’UTR

Host
Horse
Rodent
Bat
Human
Simian

0.1
5.3.4 Clinical characterisation of EPgV infection

The 12 EPgV viraemic horses comprised five horses (4 females, 1 gelding; age 5-9 years) from a single premise in Scotland and eight horses (6 mares, 1 gelding, 1 unknown, aged 5-29 years with 1 unknown age) from separate individual premises (Table 5.4).

<table>
<thead>
<tr>
<th>Horse</th>
<th>Sample Number</th>
<th>Collection Date</th>
<th>Location</th>
<th>Age, yr. /Sex</th>
<th>Pathology</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>EF_381/97</td>
<td>April-July1997</td>
<td>Scotland</td>
<td>5/F</td>
<td>None</td>
</tr>
<tr>
<td>2</td>
<td>EF_382/97</td>
<td>April-July1997</td>
<td>Scotland</td>
<td>8/F</td>
<td>None</td>
</tr>
<tr>
<td>3</td>
<td>EF_383/97</td>
<td>April-July1997</td>
<td>Scotland</td>
<td>6/F</td>
<td>None</td>
</tr>
<tr>
<td>4</td>
<td>EF_384/97</td>
<td>April-July1997</td>
<td>Scotland</td>
<td>8/F</td>
<td>None</td>
</tr>
<tr>
<td>5</td>
<td>EF_385/97</td>
<td>April-July1997</td>
<td>Scotland</td>
<td>6/F</td>
<td>None</td>
</tr>
<tr>
<td>6</td>
<td>EF_610/12</td>
<td>May 2012</td>
<td>Hampshire</td>
<td>29/F</td>
<td>Elevated GGT and AP</td>
</tr>
<tr>
<td>7</td>
<td>EF_60712</td>
<td>April 2012</td>
<td>Hampshire</td>
<td>16/MN</td>
<td>Anaemic, elevated GGT/GLDH/ Bile acids</td>
</tr>
<tr>
<td>8</td>
<td>EF_60412</td>
<td>May 2012</td>
<td>Hampshire</td>
<td>NA¹</td>
<td>Elevated GGT</td>
</tr>
<tr>
<td>9</td>
<td>EF.FR/7</td>
<td>April 2013</td>
<td>France</td>
<td>1/F</td>
<td>Elevated GGT &amp; AST²</td>
</tr>
<tr>
<td>10</td>
<td>EF.FR/11</td>
<td>April 2013</td>
<td>France</td>
<td>1/F</td>
<td>Elevated GGT &amp; AST²</td>
</tr>
<tr>
<td>11</td>
<td>EF.FR/13</td>
<td>April 2013</td>
<td>France</td>
<td>1/M</td>
<td>Elevated GGT &amp; AST²</td>
</tr>
<tr>
<td>12</td>
<td>EF.FR/15</td>
<td>April 2013</td>
<td>France</td>
<td>1/F</td>
<td>Elevated GGT &amp; AST²</td>
</tr>
</tbody>
</table>

¹NA refers to data not available
²Pasture toxins linked as causative agent
³GGT; gamma-glutamyl transpepdidase, AST; aspartate aminotransferase, AP; Alkaline phosphatase.

Based on the numbering of EPgV NC020902 nucleotide fragment 9251-9540 was used.
Seven of twelve EPgV viraemic horses were from the hepatopathy group; 4 of these had only serum biochemical evidence but no clinical indication of hepatopathy, while clinical information for the other 3 was unavailable (Horse 6, 7, 8). Of these 3 cases horse 7 had documented elevations in GGT, GLDH and bile acids and at the time of sampling was being treated for anaemia (Table 5.4, 5.5), however no historical or follow up data was available. Five out of 12 EPgV viraemic horses were from the control horse group, being clinically healthy and having no prior indication of hepatopathy. However subsequent testing of the viraemic control horses revealed that 2/5 had serum biochemical evidence of mild hepatopathy. Thus overall 9/12 EPgV viraemic horses had serum biochemical evidence of hepatopathy, namely elevation in GGT (8/12), GLDH (3/12) and/or bile acids (n=2)(Table 5.5). Based on the criteria outlined in section 1.6.1 elevated GLDH in combination with elevations in GGT and bile acids are considered as indicative of underlying hepatopathy for the purposes of this study. All control horses for which this criteria was observed (horse 2 and 3) displayed normalised or declining levels of GLDH and/ GGT or bile acids at the time of final sampling, however no further samples were available to confirm the trend detected. In the hepatopathy group no further clinical data was available for horse 7 but it appeared to display signs of hepatopathy, while horses 9-12 continued to display elevated GGT and AST but with liver biopsies taken in France detecting no liver damage or inflammation.

Viral load was determined by qPCR against NS3 transcript standard and ranged from $4.09 \times 10^5$ to $1.98 \times 10^9$ RNA copies/ml (Table 5.4). Repeat sampling of 5 of the initially viraemic horses demonstrated persistent infections were maintained over 4 month period although with declining viral loads (Table 5.4). There was little evidence for hepatopathy maintained over that period with liver enzyme levels largely within the normal range with the exception of mildly elevated GGT/GLDH/bile acids in two cases (Table 5.4).
TABLE 5.5: Viral load analysis and liver enzyme data for EPgV RNA positive horses. EPgV viraemic samples were collected between 1997 and 2013. Viral load analysis of serial samples (collected in 1997 over 4 month period between end of April and July 1997) taken from 5 RNA positive horses on a single farm indicates a persistent infection. Liver enzyme data from all viraemic horses did not suggest any underlying liver disease and with few exceptions GGT, GLDH and bile acids remained with the normal range.

<table>
<thead>
<tr>
<th>Horse</th>
<th>Group</th>
<th>Sample Number</th>
<th>Viral Load(^1)</th>
<th>GGT(^2) (≤42U/L)</th>
<th>GLDH(^2) (≤12U/L)</th>
<th>Bile Acids(^2) (0-12 umol/l)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Control</td>
<td>EF_381/97</td>
<td>1.46 x 10(^8)</td>
<td>0</td>
<td>4</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2.79 x 10(^7)</td>
<td>2</td>
<td>4</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2.34 x 10(^6)</td>
<td>9</td>
<td>7</td>
<td>4</td>
</tr>
<tr>
<td>2</td>
<td>Control</td>
<td>EF_382/97</td>
<td>4.88 x 10(^6)</td>
<td>18</td>
<td>12</td>
<td>15.4</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>3.19 x 10(^8)</td>
<td>17</td>
<td>12.5</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2.91 x 10(^6)</td>
<td>23</td>
<td>9</td>
<td>10</td>
</tr>
<tr>
<td>3</td>
<td>Control</td>
<td>EF_383/97</td>
<td>1.98 x 10(^9)</td>
<td>44</td>
<td>23</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>3.19 x 10(^8)</td>
<td>42</td>
<td>27</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2.79 x 10(^7)</td>
<td>40</td>
<td>21</td>
<td>10.6</td>
</tr>
<tr>
<td>4</td>
<td>Control</td>
<td>EF_384/97</td>
<td>3.06 x 10(^8)</td>
<td>21</td>
<td>4</td>
<td>6.8</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1.26 x 10(^6)</td>
<td>23</td>
<td>2</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1.16 x 10(^6)</td>
<td>23</td>
<td>2</td>
<td>6</td>
</tr>
<tr>
<td>5</td>
<td>Control</td>
<td>EF_385/97</td>
<td>1.59 x 10(^9)</td>
<td>34</td>
<td>9</td>
<td>10.2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2.25 x 10(^7)</td>
<td>30</td>
<td>10</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>4.09 x 10(^5)</td>
<td>34</td>
<td>8</td>
<td>10</td>
</tr>
<tr>
<td>6</td>
<td>Hepatopathy</td>
<td>EF_610/12</td>
<td>1.26 x 10(^9)</td>
<td>63</td>
<td>8</td>
<td>6</td>
</tr>
<tr>
<td>7</td>
<td>Hepatopathy</td>
<td>EF_607/12</td>
<td>1.15 x 10(^9)</td>
<td>156</td>
<td>50</td>
<td>28</td>
</tr>
<tr>
<td>8</td>
<td>Hepatopathy</td>
<td>EF_604/12</td>
<td>1.26 x 10(^9)</td>
<td>136</td>
<td>0</td>
<td>6</td>
</tr>
<tr>
<td>9</td>
<td>Hepatopathy</td>
<td>EF_FR/7</td>
<td>Below Detection</td>
<td>749</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>10</td>
<td>Hepatopathy</td>
<td>EF_FR/11</td>
<td>Limit</td>
<td>243</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>11</td>
<td>Hepatopathy</td>
<td>EF_FR/13</td>
<td>Below Detection</td>
<td>243</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>12</td>
<td>Hepatopathy</td>
<td>EF_FR/15</td>
<td>120</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
</tbody>
</table>

\(^1\) Viral loads determined by qRT-PCR
\(^2\) Elevated values are marked in bold.
5.3.5 Frequency of TDAV infections

Previously published primers targeting the conserved NS3 helicase and NS5B RdRp of TDAV (Chandriani et al., 2013) were used to screen all 328 horses and 100 donkeys. Of these samples a total of 111 horses and 100 donkeys had displayed clinical indications of underlying hepatic disease, including significantly elevated liver enzymes or evidence of hepatopathy by liver biopsy. All samples were PCR negative using the previously published primer sets. TDAV-NS3 specific primers designed as part of this study and tested against TDAV-NS3 transcript control also failed to detect the presence of viral RNA by nested PCR in all equine samples (Table 5.1). It remains to be determined whether TDAV is present in equine blood products or in horses or other species outside the United States.

5.3.6 Secondary structure of EPgV 5’UTR

The 5’UTR of EPgV is predicted to be 615 nt long, similar in length to those of other pegiviruses (Stapleton et al., 2011, Simons et al., 1996). To improve accuracy of RNA structure predictions, complete or partial 5’UTR sequences from 12 study isolates in addition to C35 (US equine isolate) and previously published sequences NC020902, KC410873 were included in the analysis. PFOLD, a stochastic context-free grammar method for identifying phylogenetically conserved co-variant sites supporting an RNA structure model (Knudsen, 2003), STRUCTUREDIST in the SSE package that identifies conserved paired and unpaired bases in minimum energy folds (Simmonds, 2012), and ALIFOLD (Gruber et al., 2008a) a combined minimum energy / covariant site detection method were implemented to generate structural predictions of the 5’UTR and results compared to determine consensus proposal.

Comparison of all three methods predicted moderate levels of consistency between predicted structures. Between position 200 and 520 of the ALIFOLD prediction, and beyond position 290 of the PFOLD model, no consistent structural predications are
generated (Appendix B). By combining PFOLD, ALIFOLD and StructureDist models (Figure 5.3) a consensus structure was predicted (Figure 5.4). The consensus prediction is a preliminary model as comparison with the 5’UTR of other known pegiviruses was not possible, due to the level of sequence diversity observed and the predicted structural differences between EPgV, HPgV and SPgV (Simons et al., 1996, Kapoor et al., 2013a). Furthermore PFOLD, ALIFOLD and StructureDist cannot predict tertiary structure present in pseudoknots, which can be observed in some viral 5’UTRs (Moes and Wirth, 2007, Paillart et al., 2002).
Structural analysis, epidemiology and clinical characteristics of EPgV infection
Epidemiology, Clinical Presentations and structural analysis of 5’UTR

FIGURE 5.3: StructureDist Dot Plot predicting secondary structural elements in the 5’UTR of EPgV. Presented here are Dot Plot images (A-D) of overlapping fragments of the EPgV 5’UTR generated using the StructureDist program from SSE V1.1. These images predict and plot clearly the presence of stem-loop structures throughout the genome. Colour coding represents the likelihood of base pairing and the likelihood that each stem-loop structure actually occurs. Genome position proceeds from left to right and bottom to top of each image, with each base-pairing presented as a set of [x,y] coordinates.
FIGURE 5.4: Consensus model for the secondary structure of EPgV 5’UTR using combined PFOLD, ALIFOLD and Structure Dist programs. By combining models generated by all three programs the EPgV 5’UTR was predicted to contain a Type 1 IRES with 2 large stem-loops D and G,
several smaller ones; A, B, C, J, K positioned 5’ and 3’ to them, and an intermediate stem-loop F a
distinctive GNRA motif, PPT tract, followed by a base paired AUG start codon and subsequent
unpaired start codon at nucleotide position 615.

**TABLE 5.6: Identified sites of covariance and semi-variance, insertions, deletions, disruptive
nucleotide pairings and unpaired sites in the preliminary modelling of the 5’UTR of EPgV.** All sites
detailed below are depicted in the modelling of the 5’UTR (Figure 5.4). Covariant sites and semi-
variant substitutions are marked in purple and pink respectively and occur to maintain base-pairing.
Regions of insertions and deletions are also predicted to occur at unpaired regions of the 5’UTR while
disruptive pairings or bases are minimal in the outlined model.

<table>
<thead>
<tr>
<th><strong>Covariant</strong></th>
<th><strong>Unpaired</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Position</strong></td>
<td><strong>Nucleotides</strong></td>
</tr>
<tr>
<td>100</td>
<td>A/G → T/C</td>
</tr>
<tr>
<td>112</td>
<td>G/C → C/G</td>
</tr>
<tr>
<td>220</td>
<td>G/A → C/T</td>
</tr>
<tr>
<td>223</td>
<td>A/G → T/C</td>
</tr>
<tr>
<td>224</td>
<td>G/A → T/C</td>
</tr>
<tr>
<td>355</td>
<td>C/A → G/T</td>
</tr>
<tr>
<td>475</td>
<td>C/T → G/A</td>
</tr>
<tr>
<td>476</td>
<td>C/T → G/A</td>
</tr>
<tr>
<td>479</td>
<td>G/C → C/T/G</td>
</tr>
<tr>
<td>481</td>
<td>G/A → C/T</td>
</tr>
<tr>
<td>482</td>
<td>A/G → T/C</td>
</tr>
<tr>
<td>589</td>
<td>T/G → G/T</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Semi-variant</strong></td>
<td><strong>Insertion</strong></td>
</tr>
<tr>
<td><strong>Position</strong></td>
<td><strong>Nucleotides</strong></td>
</tr>
<tr>
<td>53</td>
<td>C/T</td>
</tr>
<tr>
<td>159</td>
<td>G/A</td>
</tr>
<tr>
<td>197</td>
<td>T/G</td>
</tr>
<tr>
<td>251</td>
<td>G/A</td>
</tr>
<tr>
<td>255</td>
<td>T/C</td>
</tr>
<tr>
<td>256</td>
<td>C/T</td>
</tr>
<tr>
<td>278</td>
<td>A/G</td>
</tr>
<tr>
<td>312</td>
<td>T/C</td>
</tr>
<tr>
<td>328</td>
<td>A/T/G</td>
</tr>
<tr>
<td>329</td>
<td>C/T</td>
</tr>
<tr>
<td>403</td>
<td>C/T</td>
</tr>
<tr>
<td><strong>Disruptive</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Position</strong></td>
<td><strong>Nucleotides</strong></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>
5.4 Discussion

Infection with HCV and its genetically related GBV-B virus was, until recently, thought to be confined to humans and experimentally infected New World primates respectively. The detection of CHV and NPHV in dogs and horses respectively provided the first indication of a much wider host range for hepacivirus infection (Burbelo et al., 2012b, Lyons et al., 2012a, Kapoor et al., 2011). More recently, detection and characterisation of a much wider and diverse range of hepacviruses was reported in bats, rodents and an Old World primate (Kapoor et al., 2013a, Kapoor et al., 2013b, Drexler et al., 2013a, Quan et al., 2013, Lauck et al., 2013). These new discoveries fundamentally revise our knowledge of viral diversity and host ranges of viruses in this genus, their epidemiology and pathogenesis. The discovery of a rodent homolog for HCV paves the way towards a small animal model for HCV infection, with much wider implications for improving our understanding of disease progression, pathogeneses and ultimately the potential for novel treatments.

5.4.1 Active hepacivirus infection in domestic horses in the UK

Initial screening for NPHV infection discussed in Chapter 4 detected NPHV RNA in 3/142 horses. As part of this expanded study an additional 482 serum or plasma samples were screened for the presence of NPHV and EPgV RNA. NPHV RNA was not detected in any further horses outside of those previously reported (3/328) (Lyons et al., 2012a) (Table 5.2). Serial sampling of one study horse previously documented declining viral loads and the clearance of RNA over a 14 month period, while remaining seropositive (Horse 3- Table 4.5), (Chapter 4 and 6).

5.4.2 Frequency of pegivirus infections in the domestic horse population

Active pegivirus infection was detected in 12/328 (3.8%) horses, while all 100 donkeys were RNA negative (Table 5.2). Detection frequency in this study was below that reported in the United States (but not significantly so p=0.061) where a study of
74 horses reported 7 RNA positives, a frequency of 9.5% (Kapoor et al., 2013a). These findings are comparable with published data which estimates infection frequencies in the human population of \( \approx 5\% \) in the developed world with upwards of 20% viraemia recorded in some developing countries (Polgreen et al., 2003, Mohr and Stapleton, 2009, Pavlova et al., 1999), while preliminary studies in horses, rodents and bats have observed frequencies estimated at between 3.9-5\% (Quan et al., 2013, Kapoor et al., 2013a, Kapoor et al., 2013b). The failure to detect any active pegivirus infections in donkeys does not rule out the possibility that a species specific variant is circulating in the population, as has been observed for other members of this genus. To date studies indicate a narrow host range for these viruses, with HPgV being found only in humans and chimpanzees, SPgV being found in New World monkeys, and EPGV, BPgV, and RPgV being detected only in horses, bats, and rodents, respectively (Stapleton et al., 2011, Quan et al., 2013). Comparable with these studies phylogenetic analysis presented here of isolates of the Pegivirus genus in the NS3 and NS5B identifies distinct species-specific clades and the two previously identified lineages of pegiviruses infecting rodents are more closely related to each other than other mammalian species, suggestive of virus-host co-speciation. The hypothesis that pegiviruses are species specific and have codiverged with mammalian evolution can only be confirmed or disputed by further studies into pegivirus infection in mammals.

TDAV was recently identified as potential aetiological agent responsible for acute equine serum hepatitis (Chandriani et al., 2013). Eight out of 22 horses administered a botulinum antitoxin developed symptoms of hepatitis in the US study were attributed to TDAV, sequenced from the antitoxin by deep sequencing. Screening for TDAV here in the UK failed to detect the presence of any active infection in UK horses, including in horses presenting with biochemically documented hepatopathy (Table 5.2). Based on the study by Chandriani et al. RNA was undetectable in animals
sampled a year later, therefore it is possible that the design of a serology assay to
detect past infection will identify TDAV antibodies in horses in the UK.

5.4.3 **Phylogenetic analysis determines species specific clade of EPgV**

Homologs to HPgV have been isolated from primates (SPgV), rodents (RPgV), bats
(BPgV) and horses (EPgV) (Kapoor et al., 2013a, Kapoor et al., 2013b, Quan et al.,
2013, Bukh and Apgar, 1997, Epstein et al., 2010). All sequences isolated from this
study were genetically distinct from those previously characterised (Kapoor et al.,
2013a). Additionally all isolates were distinct from a second equine pegivirus, TDAV,
isolated from horses suffering from acute serum hepatitis associated with the
administration of an antitoxin (Chandriani et al., 2013). Unlike other currently
characterised naturally occurring pegivirus infections, this is the first to be associated
with hepatic disease, although future studies are required to confirm the association.
The diversity of isolates observed in rodents and particularly bats implies that these
viruses have been circulating in the population for an exceedingly long time and the
presence of several rodent and bat pegivirus clades (Figure 5.1 and 5.2) implies the
long-term evolution of these viruses with their hosts.

To date studies indicate a narrow host range for each of these viruses, with HPgV
being found only in chimpanzees and humans, SPgV detected only in New World
monkeys, and EPgV, RPgV and BPgV detected only in horses, rodents and bats
respectively. In this study we further failed to detect EPgV in any other species
screened. These findings in addition to the distinct species-specific clades
characterised in Figure 5.1 and 5.2 and also in previously published studies are in
agreement with the hypothesis that pegiviruses infecting horses and other mammalian
species are species-species and have codiverged with the evolution of mammals
(Sharp and Simmonds, 2011, Kapoor et al., 2013b, Kapoor et al., 2013a, Simmonds,
2001a). Supporting this hypothesis is the observation that the different human
genotypes are prevalent in different populations and may have emerged during the evolution of humans and subsequent journey out of Africa (Tanaka et al., 1998d, Tanaka et al., 1998c).

5.4.4 Analysis of correlation between NPHV and EPgV viraemia and hepatopathy in horses

The scope of this study allowed for some analysis into the pathogenesis of NPHV and EPgV infection. The classification and re-naming of members of the Pegivirus genus was partly in response to the accumulating failure to demonstrate any link between human infection with HPgV and hepatitis, making description such as “hepatitis G virus” inappropriate (Stapleton et al., 2011; Takikawa et al., 2010). Data from this study facilitated some preliminary analysis of potential associations of active NPHV and EPgV infection with hepatopathy in horses.

Serum biochemical evidence of hepatopathy was identified in 1/3 horses with NPHV viraemia and 9/12 horses with EPgV viraemia, but further study is required to determine whether the hepatopathy was attributable to NPHV and EPgV infections or to another of the many causes of equine hepatopathy. The prevalence of NPHV and EPgV viraemia were not significantly different in horses with hepatopathy versus the control group, suggesting that these viruses are unlikely to represent a common cause of hepatopathy in the study population. In contrast, a previous study detected an increased frequency of EPgV viraemia (25%, 3/12) in horses with serum biochemical evidence of hepatopathy compared to healthy controls (6.4%, 4/62) but also concluded that their study was insufficient to determine the health relevance of EPgV infection (Kapoor et al., 2013b). However, combining data from the current study and this previous study provides statistically significant evidence for an association between hepatopathy and EPgV viraemia; with combined infection frequencies of 8.1% (10/123) in horses with hepatopathy compared to 3.2% (9/279) in control horses (Fisher’s exact test; p=0.042). This significant association and the prior detection of
viral RNA in the liver and lymph of infected horses (Kapoor et al., 2013b) provides preliminary data suggesting that EPgV may be a cause or contributory factor in equine hepatopathy. However the observation that all 3 NPHV and all 6 EPgV viraemic horses for which clinical data were available were clinically healthy contrasts with HCV infections in humans which causes significant liver damage even in early stages of infection (Alter et al., 1995, Engel et al., 2007, Fierer et al., 2008). Further study is required to clarify potential disease associations with NPHV and EPgV infections in horses.

Evidence of persistent EPgV infection as seen with SPgV and HPgV was apparent in serial equine samples screened over 4 month period in 5 horses from the same estate (Table 5.5), who also remained antibody positive throughout the course of viraemia (Chapter 6-Table 6.9). Infection with SPgV has been frequently documented to result in persistent viraemia (Simons et al., 1995b), in contrast experimental infection of tamarins with GBV-B results in viral clearance within 6 months (Bukh et al., 2001b, Jacob et al., 2004, Simons et al., 1995b). In the context of HPgV infection, immune competent individuals generally clear infection within 2 years (Tanaka et al., 1998a, Berg et al., 1999). Unfortunately samples were unavailable from the 5 horses outside of this 4 month period in 1997, although fluctuating viral loads (Table 5.5) indicate trend towards ultimate clearance of viral RNA, as documented for other pegivirus species, SPgV, HPgV and TDAV but as yet not documented for RPgV and BPgV. Antibodies to EPgV were confirmed in this study in 10/12 of the RNA positive cases and appeared to remain elevated or elevating in the serial samples taken from horses 1-5 (Chapter 6). Over the same 4-month period a general declining trend in viral load was observed. In contrast to HCV which elicits antibodies to several viral proteins during viraemia that persist throughout infection (Baumert et al., 2000), antibodies to HPgV are not generally detectable during active viraemia, although some studies report the detection of anti-HPgV antibodies (Gomara et al., 2010, Pilot-Matias et al.,
1996b, Schwarze-Zander et al., 2006, Tan et al., 1999, Van der Bij et al., 2005). In addition there are no documented studies recording the detection of SPgV antibodies, although this may be a consequence of lack of suitable reagents for detection (Schlauder et al., 1995a). Subsequent studies on the seroprevalence of EPgV globally will provide more insight into the exposure rates among horses worldwide.

5.4.5 Absence of hepaci-like viruses in NHPs

Finally as part of this study all NHPs were screened for homologs of HCV and NPHV. Samples originated from wild or rescued chimpanzees and gorillas in Cameroon, a region of endemic HCV infection in Africa. Comparable to previously published studies surveying hepatitis infections in NHP’s (Makuwa et al., 2003, Makuwa et al., 2006), all samples were PCR negative. GBV-B a member of the Hepacivirus genus related to but divergent from HCV, causes hepatitis in lab infected tamarins, but its ultimate origins remain unknown and the virus has not been isolated infecting New World monkeys in the wild. Based on this and the recent detection of hepaciviruses in rodents (RHV) (Drexler et al., 2013a, Kapoor et al., 2013b) and bats (BHV) (Quan et al., 2013) which are highly divergent from HCV and NPHV it is not possible to rule out the possibility that a highly divergent hepacivirus remains to be detected by PCR or deep sequencing of NHP serum. Recent detection and characterisation of a novel hepacivirus in New World primates, displaying an uncharacteristically long NS5A compared to other members of the Flaviviridae shed new light on the evolutionary history and viral and host diversity associated with members of the Flaviviridae family (Lauck et al., 2013). Three isolates of the virus termed quereza hepacivirus (GHV) from colobus monkeys in Uganda shared between 98-85% nucleotide sequence similarity. However, GHV shared limited nucleotide sequence similarity of between 43-50%, to other members of the Hepacivirus genus; HCV, NHPV, RHV, GBV-B and BHV respectively (Lauck et al., 2013). As a consequence of these findings, the failure to detect hepacivirus variants amongst the NHPs screened in this study may reflect the
presence of divergent viral variants as yet undetectable by the PCR methodologies applied, due to the lack of appropriate genomic sequence data.

### 5.4.6 Secondary structure prediction of 5’UTR and IRES elements of EPgV

The proposed consensus structure for the EPgV 5’UTR comprises 2 large stem-loops (labelled D and G), upstream and downstream of these stem-loops 5 additional smaller ones (A, B, C, J, K), and a final intermediate stem-loop F. These structures are larger but similarly positioned to stem-loops II and IV described in other pegiviruses; HPgV and SPgV (Simons et al., 1996) and stem-loop H is positioned similarly to SL-IVB (Figure 5.4). Comparison of sequence data from variants of EPgV strongly supported the predicted model. Where sequence variability occurred, most nucleotide insertions/deletions and variable sites were confined to regions predicted as being unpaired (shaded green) but for two exceptions (Table 5.6). Variable sites that were predicted to occur in paired regions display co-variant or semi-covariant compensatory substitutions in order to maintain pairing (shaded blue and red). Several features of the 3’ terminal half of the EPgV 5’UTR are further consistent with an IRES structure. Between nucleotide positions 574-590 an unpaired pyrimidine tract (PPT) is predicted, and additionally, stem-loop H positioned upstream of the PPT contains a terminal GNRA tetraloop found in most virus IRES structures (Pilipenko et al., 1989, Pilipenko et al., 1994, Gan et al., 1998, Kaminski et al., 1994, Nicholson et al., 1991, Kohara et al., 1991). Indeed, the arrangement of stem-loops G and H followed by the PPT, a cryptic base-paired start codon in stem-loop K and an additional unpaired region and finally the likely authentic AUG start codon is remarkably similar to the structure of type 1 IRESs of enteroviruses (in the virus family Picornaviridae). Nonetheless, future physical mapping of the 5’UTR and functional characterization are required future to confirm this putative structural homology.
Recent studies on the 5’ UTR of individual RHV isolates indicate the presence of both type 4 and type 3 IRES elements comparable to predicted pegivirus and hepacivirus 5’UTR structures, respectively (Drexler et al., 2013a). The prediction of a pegi-like IRES element in the 5’ UTR of RHV provides suggestive evidence for potential recombination between IRES elements. Further supporting evidence for this is provided here and in previous publications suggesting the presence of type 1 picorna-like and type 4 pegi-like IRES elements in the 5’ UTR of the EPgV genome (Kapoor et al., 2013a).

5.4.7 Conclusion

Identification of equine homologs of HCV and HPgV circulating in horses marked the beginning of studies into the detection and characterisation of divergent hepaciviruses and pegiviruses circulating in broad range of mammals; rodents, bat, and NHPs across the globe. Findings presented here suggest NPHV is present in the UK equine population and EPgV infections are present in horses in the UK and France with no determined disease associations for either virus.

All EPgV sequence isolates reported here are distinct from those previously reported, in addition to TDAV; associated with acute serum hepatitis in horses. In contrast to TDAV, findings in this study detected no evidence of any negative health effects upon the horses of EPgV infection despite the high levels of EPgV viraemia detected. As no liver biopsies or lymph fluid was available for testing the exact tissue tropism of the virus could not be conclusively determined. The recent detection of a novel and highly divergent hepacivirus in an OWM highlights the possibility that divergent and as yet undetected hepaci-and pegi-like viruses are present and circulating in the wild amongst NHPs and other species (Lauck et al., 2013).

Continued PCR of other mammalian taxa across a range of geographical locations, including areas of endemic HCV and HPgV infection is required to address these gaps
in our understanding of the evolution of hepaciviruses and pegiviruses. The design of methodologies for the accurate and sensitive detection of past hepacivirus and pegivirus infection by detection of serum antibodies will greatly improve our understanding of disease pathogenesis, host range, host immune responses, and virus dissemination and transmission routes.
Chapter 6
Prevalence of NPHV and EPGV antibodies across species

6.1 Introduction

6.1.1 Diagnostic methods for the detection of hepacivirus and pegivirus exposure in mammals

Hepacivirus and pegivirus infections affect approximately 3% and 5% of the world’s population respectively (Sulaiman et al., 1995, Shepard et al., 2005, Alter et al., 1999, Desenclos, 2000, Frank et al., 2000). HCV is a leading cause of chronic liver disease, hepatocellular carcinoma and liver cirrhosis, in contrast to HPgV, which is a lymphotropic virus of unknown pathogenesis in humans (Laskus et al., 1997, Radkowski et al., 1999, Tucker et al., 2000). Essential in the treatment of HCV is correct and effective diagnosis, leading to the provision of the most efficient treatment regimens (Bornschlegel et al., 2013). Currently diagnosis and screening for on-going and historical HCV infection is based on the combined detection of anti-HCV antibodies and a confirmatory PCR to detect HCV RNA (Bornschlegel et al., 2013). Detection of serological markers of HCV infection currently applies assays targeting the highly immunogenic regions of the virus core, NS3 and NS5 (Alter, 1994, Courouce et al., 1995, Courouce et al., 1994, Vallari et al., 1992, Pawlotsky et al., 1994).

Recently the importance applying such detection methods to other species in an attempt to greater understand the origins of HCV was highlighted by the discovery by PCR and deep sequencing of a diverse range hepacivirus in rodent (RHV), bats (BHV), dogs (CHV), and OWMs (GHV) (Drexler et al., 2013a, Kapoor et al., 2011, Kapoor et al., 2013b, Lauck et al., 2013, Quan et al., 2013). Furthermore the application of novel serology based methods to detect past CHV infection in a range
of mammal species detected CHV NS3 helicase antibodies in horses that subsequently led to the isolation and characterisation of NPHV (Burbelo et al., 2012b). The novel LIPS assay uses antigens expressed in cells as fused Renilla Luciferase (Ruc)-antigen recombinants and is recognised for its sensitivity and ease of production (Burbelo et al., 2009, Burbelo et al., 2012a, Ramanathan et al., 2008, Burbelo et al., 2005). An indirect immunoflorescent assay (IFA) was also recently applied using HCV-infected HuH7-cells (strain JC1) or replicon JFH1-transfected cells to screen rodent and bat samples for anti-HCV antibodies (Drexler et al., 2013a). 180 bat and 95 rodent serum samples were tested against immunofluorescent slides containing full recombinant HCV and 7.2% (13/180) of bats were seroreactive, but no rodents. Western blots confirmed anti-HCV antibodies in 10/13 bats tested against HCV-NS3, helicase, NS5, NS4 and core proteins. These findings prior to the publication of RHV detection by PCR may be the result of assay non-specificity or were indeed the early indicators of diverse hepaciviruses in bats that due to the level of sequence diversity had not yet been detected by conventional PCR (Quan et al., 2013, Drexler et al., 2013a). PCR methodologies and deep sequencing technologies however led to the first detection in NHPs of a divergent hepaciv-like virus (Lauck et al., 2013), a significant step in understanding the origins of HCV which have long been postulated to originate in NHPs, as was the case for HIV-1 and HIV-2 whose origins were traced to chimpanzees (Gao et al., 1999).

Infection with HPgV, in contrast to HCV is not associated with hepatic disease and interest in the virus waned when studies failed to detect any disease associations (Batts, 1997). When studies on HPgV pathogenesis in HIV infected patients indicated an inhibitory effect on HIV replication interest was revived (Devereux et al., 1998, Hollingsworth et al., 1998, Mohr and Stapleton, 2009, Tillmann et al., 2001). Diagnosis of acute and chronic HPgV infection is based on the detection of anti-E2 antibodies in the presence or absence or viral RNA (Tacke et al., 1997a, Tacke et al., 1997b, Dille et al., 1997). Nucleic acid testing (NAT) and the
application of next generation sequencing technologies has led to the detection of pegivirus infection in addition to those infecting humans and NHPs in a range of mammals including rodents (RPgV), bats (BPgV) and horses (EPgV) (Quan et al., 2013, Epstein et al., 2010, Simons et al., 1995a, Kapoor et al., 2013a, Kapoor et al., 2013b).

6.1.2 Hepacivirus transmission between species

While the search for the origins of HCV focused on NHPs until recently due to the associated origins of HIV with chimpanzees (Gao et al., 1999), a novel canine homolog was isolated from dogs suffering from severe respiratory infections in the US (Kapoor et al., 2011). Unusually the viral load was significantly higher in bronchial lavage samples compared to liver samples (>10^7 compared to <10^3 genomic RNA copies per 2ng of total RNA). In subsequent studies CHV RNA could not be detected in any of 213 canine serum samples suggesting that active infections were not present in the UK canine population (Lyons et al., 2012a, Bexfield et al., 2014). The severity of the infection and the detrimental outcome to the dogs in the US and the subsequent detection of NPHV isolate in horses of 99% sequence identity to the CHV isolate suggests that the virus may in fact have been transmitted from horses to dogs, through as yet unknown transmission routes. This hypothesis is tentatively supported by more recent findings of extremely divergent but related hepacviruses in rodents, bats and New World monkeys. Initial phylogenetic analysis implies that bats may be an ancient natural reservoir for hepacivirus genetic diversity and may have been zoonotically transmitted to humans and to other mammalian species through the hunting and consumption of infected bushmeat. While initial studies have not detected or focused on any clinical or biological consequences of hepacivirus infection in these species, they do confirm the potential for successful cross-species transmission events. Widespread screening of canines and indeed other mammalian species and sub-species in regions of endemic HCV infection or in more widespread contact with infected animals will provide greater insight into the on-
going evolution and dissemination of these viruses which will impact upon global control of HCV.

6.1.3 Level of exposure in global population to hepacivirus and pegivirus infections as a measure of antibody reactivity

Infection with HCV and HPgV is globally distributed and approximately that 2-5% of the human population are affected. The prevalence of anti-HCV in individual population studies is used to estimate the global burden of disease. Based on historical evidence the level of exposure has been documented to be highest in Africa and Asia, while other countries of the developed world; Australia, western Europe and North America have the lowest reported levels (WHO, 1999, Lavanchy, 2009, Mohd Hanafiah et al., 2013, Shepard et al., 2005). A recent review of all available literature on the seroprevalence of HCV by The Global Burden of Disease 2010 concluded that the level of people with anti-HCV had increased from 2.3% to 2.8% or from less than 122 million to under 185 million (Mohd Hanafiah et al., 2013). In regions of endemic HCV infection like Egypt and Pakistan seroprevalence rates of ≈4% and 15% respectively are observed (Qureshi et al., 2010, Miller and Abu-Raddad, 2010, Averhoff et al., 2012). Homologs of HCV have been detected in recent years circulating in rodents, bats, horses and OWMs, although the seroprevalence has only been estimated amongst horses thus far (Burbelo et al., 2012a). The 35% level of exposure in the horses surveyed is within the range of exposure associated with HCV infection amongst IDUs and other risk groups, where studies have reported between 37-85% of participants as HCV antibody positive (Garfein et al., 1998, Garfein et al., 1996, Thomas et al., 1995, Chelleng et al., 2008)

As previously discussed (Chapter 5) the frequency of HPgV infection is difficult to determine but population based studies have estimated ≈3% of healthy blood donors in the developed world are viraemic and a further ≈9% have anti-E2 antibodies (Blair et al., 1998, Gutierrez et al., 1997, Pilot-Matias et al., 1996a, Tacke et al., 1997a).
Examination of risk groups such as those with sexually transmitted or blood borne diseases detect increased levels of HPgV exposure (Scallan et al., 1998b). Antibodies have been detected in 46% of HIV infected homosexual men in a single study (Williams et al., 2004). Levels of viraemia have been estimated for BPgV, RPgV and EPgV at between 3.9-5% as previously discussed (Chapter 5), although the total exposure rate has not been estimated amongst these species by serology based assays methods. Additionally the course of infection and the clinical significance of pegivirus infection amongst seropositive and antibody positive rodents, bats and horses has not been thoroughly studied to date.

6.1.4 Antibody detection and the establishment of persistent hepaci- and pegivirus infections

HCV and SPgV frequently leads to persistent infections (Hoofnagle, 1997, Lauer and Walker, 2001, Simons et al., 1995b), while 80% of healthy individuals spontaneously clear HPgV viraemia (Stark et al., 1997, Tanaka et al., 1998b, Gutierrez et al., 1997), although viraemia has been documented to persist for decades (Alter et al., 1997b). Antibodies to several HCV viral proteins remain detectable during viraemia (Baumert et al., 2000) in contrast to the large proportion of HPgV infections which are cleared within 2 years by immune competent individuals (Berg et al., 1999, Tanaka et al., 1998a), after which E2 antibodies are detectable and serve as a marker of past infection (Gutierrez et al., 1997, McLinden et al., 2006, Tacke et al., 1997a, Tanaka et al., 1998a, Tan et al., 1999). There are some studies however, which report the detection of anti HPgV peptide reactivity during viraemia (Gomara et al., 2010, Pilot-Matias et al., 1996a). Antibodies have not been detected against SPgV but this may be associated with the lack of availability of appropriate reagents (Simons et al., 1995b). Tamarins usually clear GBV-B infection within 6 months and no persistent infections are established (Bukh et al., 2001b, Jacob et al., 2004, Simons et al., 1995b). To date there is no serological data available to determine the
course of EPgV, BPgV or RPgV infection in these species, and the point at which viraemia peeks or seroconversion is established.

In most HCV infected individuals, anti-HCV antibodies appear 7 to 8 weeks after exposure (Bowen and Walker, 2005), this can be delayed slightly in cases of transfusion acquired infections (Glynn et al., 2005). Between 80-97% of patients develop antibodies to HCV by 15 weeks to 6 months (CDC, 1998). Seroconversion is rarely delayed beyond one-year post infection (Beld et al., 1999), while patients with HIV infections may not actually seroconvert (George et al., 2002). What remains unclear is the role of antibody production in the control of HCV. Antibodies are frequently detected in individuals who clear viraemia, but over time antibody titres decline, and there may be no serological evidence of infection in a large number of infected cases. Antibodies do not appear to confer protection against re-infection, and the immune system of some individuals appears to control viraemia in the absence of seroconversion (Bowen and Walker, 2005). On the other hand a strong CD4 and CD8 HCV specific T-cell response generally correlates with recovery from acute infection and often coincides with observed elevations in the levels aminotransferase (Bowen and Walker, 2005).

Research presented within this chapter extends data currently available on the seroprevalence and infection frequencies of NPHV and EPgV. Through the use of novel NPHV and EPgV ELISA assays we examine the relationship between antibody detection and viral load during and post viraemia in an attempt to better understand the dynamics of these infections and their correlation if any with their human homologs. This study confirms the detection of the first NPHV seropositive dog, which supports the initial detection of CHV RNA in dogs in kennels in US, suggesting the potential for zoonotic transmission of the virus between species. Through the amalgamation of studies into active infection presented in Chapter 5
with those of antibody detection presented within this chapter it is possible to improve our understanding into NPVH disease progression.

Similarly in order to further our comprehension of EPgV, novel assays presented in this chapter were designed to determine the level of past exposure in horses and examine the correlation if any with liver disease through analysis of liver enzyme levels and medical records. Through examination of serial RNA positive samples (Chapter 5) the relationship between antibody detection and the presence or absence of viral RNA can be examined over the course of infection and used as a potential marker to indicate the presence or absence of persistent infection in the host.

6.2 Materials and Methods

6.2.1 Sample selection

Serum or plasma samples (n=1197) were collected from 362 humans (71 HCV positive male IDUs, 291 HCV negative men from Cameroon), 163 NHPs (Cameroon), 328 horses (Scotland, England and France), 100 donkeys (England), 113 dogs (Scotland) and 131 cats (Scotland). Samples from humans and NHPs were collected in Cameroon by Metabiota (formerly Global Viral Forecasting Initiative) staff in Cameroon and included 11 gorillas, 62 chimpanzees and 91 Old World Monkeys (OWM) as previously outlined (Lavoie et al., 2012; Lyons et al., 2012b; Sharp et al., 2010). Samples from horses, donkeys, dogs and cats were archived excess diagnostic samples collected at the Royal (Dick) School of Veterinary Studies, Edinburgh, Liphook Equine Hospital Laboratory, Hampshire and The Donkey Sanctuary, Sidmouth. Animals were of mixed breed, age and sexes. Donkeys, dogs and cats had a wide range of diseases that prompted diagnostic sampling. Horses were categorised into a hepatopathy group (n=111) and a control group (n=217). The hepatopathy group comprised horses that had been investigated for suspected liver disease and which had biochemical evidence of hepatopathy, as
indicated by serum/plasma concentrations of gamma-glutamyl transferase (GGT), glutamate dehydrogenase (GLDH) and/or bile acids exceeding the laboratory reference upper level. Hepatopathy was attributed to a wide range of causes. The control group comprised horses for which there was no clinical suspicion of hepatopathy, including clinically healthy horses and horses that were sampled for investigation of a wide range of diseases excluding hepatopathy. Samples were collected between 1995 and 2013 and anonymised prior to testing.

All NHP blood samples were collected via venepuncture from 73 apes comprising 11 gorillas (Gorilla gorilla) and 62 chimpanzees (Pan troglodytes troglodytes and Pan troglodytes ellioti), and from a variety of Old World Monkey species: Cercocebus agilis (n = 7), C. torquatus (n = 2), Cercopithecus cephus (n = 3), C. erythrotis (n = 4), C. l'hoesti preussi (n = 4), C. mona (n = 9), C. nictitans (n = 3), C. pogonias (n = 1), C. tantalus (n = 3), Erythrocebus patas (n = 3), Lophocebus albigena (n = 5), Mandrillus leucophaeus (n = 20), M. sphinx (n = 9), and Papio anubis (n = 20). If not previously recovered serum was separated from whole blood by centrifugation and stored long term at -80ºC. Samples were shipped to the United Kingdom from Cameroon in compliance with UK and Cameroon laws and the Convention on International Trade in Endangered Species of Wild Fauna and Flora (CITES) (Sharp et al., 2010).

### 6.2.2 Serology

#### 6.2.2.1 PCR amplification of NS3 and Core regions

Nested PCR primers for NPHV and EPgV NS3 helicase were designed based on published sequences from, with appropriate restriction sites for sub cloning added to the inner primers (Table 6.1). PCRs were performed using GoTaq (Promega) in accordance with the manufacturers’ instructions, by using conditions outlined in Chapter 2.
6.2.2.2 Synthesis of NPHV and EPgV NS3 and core recombinant proteins

Amplicons were cloned into the pGex-2T or pET28B Vector for production of Glutathione-S-Transferase and polyhistidine fusion proteins respectively (Amersham Pharmacis Biotech, Merck Novagen). Briefly the plasmid and the inserts were digested with BamHI and EcoRI according to manufacturer’s instructions (Promega) and gel purified using PureLink Gel Extraction Kit (Invitrogen) (Chapter 2). Ligation reactions were carried out using T4 Ligase (Bioline) at insert: vector ratio of 4 and transformed into JM109 competent cells (Agilent) for propagation. Antibiotic resistance to ampicillin and gentamicin was used to select for recombinant pGex-2t and pET28B clones, respectively. Vector specific PCRs and subsequent sequencing was used to ensure all clones contained correct coding sequences. High concentrations of plasmid were purified from overnight cultures of ampicillin/gentamicin resistant colonies using Wizard® Plus SV Minipreps DNA Purification System as per manufacturer’s instructions (Promega).

**TABLE 6.1: Details of recombinant plasmids generated, and predicted molecular weights, confirmed by SDS-PAGE analysis**

<table>
<thead>
<tr>
<th>Recombinant ID</th>
<th>Vector</th>
<th>Insert</th>
<th>Tag</th>
<th>Molecular Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>NPHV_NS3_GX</td>
<td>pGex-2t</td>
<td>NS3-NPHV</td>
<td>GST</td>
<td>55kDa</td>
</tr>
<tr>
<td>NPHV_Core_GX</td>
<td>pGex-2t</td>
<td>Core_NPHV</td>
<td>GST</td>
<td>49kDa</td>
</tr>
<tr>
<td>EPgV_NS3_pET</td>
<td>pET-28B</td>
<td>NS3-EPgV</td>
<td>6xHis</td>
<td>37kDa</td>
</tr>
</tbody>
</table>

6.2.2.3 Recombinant protein expression

As outlined in section 2.2.8.4 the optimum conditions for induction of protein expression were determined based on careful examination of the level of expression detected by SDS-PAGE over a range of temperatures between 37°C and 20°C and over a period of induction between 1-24 hours. At temperature between 30-37°C and for periods of induction greater than 4 hours the level of observed protein expression in the insoluble cell lysate increased compared to marked decrease in that detected in the soluble cell lysate (Figure 6.1). To increase the solubility of the recombinant
protein the temperature of induction was reduced and protein expression was sampled from 0 hours to 18 hours. The final concentration of IPTG used for induction was also analysed at levels between 0.1mM and 1mM, however this was not found to significantly alter the level of protein expression in the soluble and insoluble cell lysates.

In the final protocol for protein expression One Shot® BL21 (DE3) pLysS Chemically Competent *E. coli* (Invitrogen) were transformed with plasmid as per manufacturer’s instructions. From overnight cultures of antibiotic resistant colonies a fresh 1:10 dilution was prepared and grown to OD 0.6 before induction at 20°C for 3 hours with 1mM IPTG. LB medium was removed by centrifugation at 4000xg for 15 minutes and cell pellets were weighed and lysed with BugBuster Master Mix (Merck Millipore) containing Roche complete ULTRA protease inhibitors, as per
manufacturer’s instructions. Cell lysate was recovered by centrifugation at 12000 xg for 15 minutes at 4°C. Soluble fractions were analysed by SDS-PAGE for the presence of recombinant protein expression (Figure 6.2) and subsequently tested by western blot using anti-6X-His (Abcam) or anti-GST (Abcam) HRP-conjugated antibodies to confirm the expression of recombinant EPgV-NS3 and NPHV-NS3/core proteins respectively (Figure 6.3).
FIGURE 6.2: SDS-PAGE analysis of unpurified E. coli lysate with recombinant proteins generated to detect NPHV and EPgV antibodies. (A) SDS-PAGE of recombinant pGex2t-NPHV-NS3 helicase. Lane 1: uninduced recombinant E. coli soluble lysate; Lanes 2: molecular weight ladder (Fisher); Lane 3-5: induced non-recombinant pGEX-2t E. coli soluble lysate; Lane 6: induced recombinant NPHV-NS3 E. coli soluble lysate, molecular weight of 55kDa. (B) SDS-PAGE of recombinant pGex2t-Core-NPHV. Lanes 1-2: induced pGex2t-Core-NPHV E. coli soluble lysate, molecular weight of 49kDa. Lane 3: Molecular weight ladder (Fisher); Lane 4: uninduced recombinant E. coli soluble lysate; Lane 5-6: induced non-recombinant pGEX-2t E. coli soluble lysate. (C) SDS-PAGE analysis of pET28B-NS3-EPgv unpurified E. coli soluble lysate. Lane 1 molecular weight ladder (Fisher); Lane 2: uninduced recombinant E. coli soluble lysate; Lane 3 and 4:
induced recombinant NS3-EPgV, molecular weight of 37kDa; Lane 5: induced non-recombinant pET28B E. coli soluble lysate.

**FIGURE 6.3:** Western blot analysis of unpurified E. coli soluble lysate with recombinant proteins generated to detect NPHV-NS3/Core and EPgV-NS3 antibodies. (A) Western blot of recombinant pGex2t-NPHV-Core using 1:25000 dilution of anti-GST HRP-conjugated antibody. Lane 1 and 5: molecular weight ladder (Invitrogen); Lane 2-3: induced recombinant E. coli soluble lysate, molecular weight of 49kDa; Lanes 4: uninduced recombinant E. coli soluble lysate; Lane 6-7: uninduced non-recombinant Lane 8: induced non-recombinant pGEX-2t E. coli soluble lysate. (B) Western blot of recombinant pGex2t-NPHV-NS3 helicase, molecular weight of 55kDa tested using 1:25000 dilution of anti-GST HRP-conjugated antibody. Lanes 1: Molecular weight ladder (Invitrogen); Lane 2: uninduced recombinant E. coli soluble lysate; Lane 3: induced non-recombinant pGex-2t E. coli
soluble lysate; Lane 4: induced recombinant E. coli soluble lysate. (C) Western blot of recombinant pET28B-NS3-EPgV unpurified E. coli soluble lysate, molecular weight of 37kDa tested using 6X anti-his HRP-conjugated antibody. Lanes 1: Molecular weight ladder (Invitrogen); Lane 2: induced recombinant E. coli soluble lysate; Lane 3: uninduced recombinant E. coli soluble lysate; Lane 4: induced non-recombinant E. coli soluble lysate.

6.2.3 Enzyme linked immunosorbent assay

All samples (n=1,156) were tested for antibodies to NPHV. Horse and donkey samples only (n=428) were also tested for antibodies to EPgV. High-bind 96-well enzyme-linked immunosorbent assay (ELISA) plates (Greiner Bio-One) were coated overnight with BugBuster cell lysates (0.5 µg of recombinant protein in 100 µL of carbonate buffer) or an equivalent volume of protein isolated from cells infected with non-recombinant pGex-2t/pET28B collected and processed in parallel with the NS3 and Core proteins of NPHV and the NS3 protein of EPgV respectively. Plates were washed with 250 µL of 1% Tween 20/PBS (vol/vol), and coated wells were blocked with 150 µL of 3% BSA/PBS (wt/vol) at room temperature for 2 h. After 1 round of washing, 100 µL of test serum samples pre-incubated with 10 X non-recombinant pGex-2t/pET28B cell lysate and diluted to a final concentration of 1:1000 in 3% BSA/PBS (wt/vol) were added to the wells and incubated for 1 h. The wells were washed 6 times with 250 µL of 1% Tween 20/PBS (vol/vol), were incubated each time for 15 min, and then were incubated for 30 min with 100 µL of horseradish peroxidase–conjugated goat anti–horse immunoglobulin G antibody (Serotec) or the appropriate species specific anti-IgG diluted 1:4000 in 2% BSA/PBS (wt/vol). After 4 rounds of washing, plates were developed by adding 70 µL of the horseradish peroxidase substrate (2,2’-azino-bis 3-ethylbenzothiazoline-6-sulfonic acid; Liquid Substrate System) to each well in accordance with the manufacturer’s protocol (Sigma). Plates were allowed to develop for 10–15 min and were read at 405 nm. The immunoreactivity of serum samples to control wells was compared and the mean control plus 3 times the standard deviation was established as cut off. Samples confirmed as NPHV NS3 antibody positive were confirmed by second ELISA
targeting NPHV Core antigen. EPgV antibody positive samples were confirmed by western blot analyses.

The antibody and serum concentrations applied in all ELISAs and outlined above were optimised prior to carrying out large-scale screening. By carrying out a design of experiment varying the final concentration of test serum (between 1:10 to 1:20000) in combination with varying with the antibody concentration (between 1:500 to 1:20000) it was determined that the best signal in the sample wells combined with the lowest level of background in the control wells was achieved using the combination of 1:1000 of sample serum with 1:4000 of the HRP conjugated species-specific antibody. Finally the percentage of BSA used for blocking and antibody dilution was also tested at concentrations between 1-5%, with no advantage observed in using concentrations exceeding 3% and this was applied in all subsequent serology screening.

6.2.4 Western blot analysis of EPgV seropositive animals

Western blots were carried out to confirm all EPgV seropositive cases in the absence of a second confirmatory ELISA for EPgV. Whole-cell lysate and enriched protein samples were analysed on 10%/16% sodium dodecyl sulphate (SDS)–polyacrylamide gel electrophoresis (PAGE) gels, depending upon the calculated protein molecular weight. The gels were either stained with Coomassie blue (Sigma) or blotted onto Protran BA 85 nitrocellulose membranes (Whatman) by use of a semidry electroblotter. Membranes and were blocked overnight in 4% Milk Powder (MP)/0.05% Tween 20/TBS (wt/vol/vol). The membranes were then incubated at room temperature for 1 h with serum obtained from horses identified as EPgV seropositive and diluted 1:2000 in 4% MP/0.05% Tween 20/TBS (wt/vol). After 6 washes in 0.05% Tween 20/TBS (vol/vol) for a total of 2 h, the membrane was incubated with horseradish peroxidase–conjugated goat anti–horse immunoglobulin G antibody (Serotec) diluted 1:35000 in 4% MP/0.05% Tween 20/TBS (wt/vol). After 4 washes in 0.05% Tween 20/TBS (vol/vol) for a total of 1 h, bound antibody
Seroprevalence of NPHV and EPgV in domestic horses was visualized using electrochemiluminescent prime western blotting detection reagent in accordance with the manufacturer’s protocol (Amersham) (Figure 6.4).

**FIGURE 6.4: Western blot detection of EPgV Anti-NS3antibodies in ELISA seropositive animals.** Serum identified as seropositive by ELISA was tested against induced and non-induced recombinant pET28B-EPgV NS3 to confirm the presence of antibodies to EPgV in the absence of secondary confirmatory ELISA. Lane 1: Positive horse serum tested on non-Induced recombinant pET28B-EPgV-NS3 *E. coli* soluble lysate; Lane 2 and 3: Positive horse serum tested against induced recombinant pET28B-EPgV NS3 *E. coli* soluble lysate (antibody binding to antigen detected at 37kDa), Lane 4: Molecular weight ladder (Invitrogen), Lane 5: Positive horse serum tested against induced non-recombinant pET28B *E. coli* soluble lysate. Anti-horse-IgG HRP was used and fluorescence measured using electrochemiluminescent prime western blotting detection reagent (Amersham).
6.3 Results

6.3.1 ELISA Screening for past exposure to NPHV across species

To investigate frequencies of current infection and past exposure to NPHV, we developed an ELISA using recombinant protein expressed from NS3-helicase domain and from the core gene as a confirmatory assay (Table 6.1). This two-stage assay was used to screen a large study group of 1,156 serum/plasma samples from horses and other mammalian species (328 horses, 100 donkeys, 116 dogs, 132 cats, 362 human, 164 non-human primates, NHP) (Table 6.2). All samples were screened in blocking concentrations of soluble non-recombinant pGex-2t E. coli lysate to minimise assay non-specificity and reactivity to NS3 / core-expressing and non-expressing controls were compared (Figure 6.5). The cut-off for the NPHV assay was set conservatively as the mean serological reactivity plus 3 SDs of unreactive samples and was used to categorize samples as anti-NPHV positive or negative by each ELISA. For the NPHV NS3 assay this corresponded to 0.128 +3(0.053) and a cut-off value of 0.287 OD, and for the core ELISA the cut-off equated to 0.136 + 3(0.048), a value of 0.280 OD (Figure 6.5).

<table>
<thead>
<tr>
<th>Species</th>
<th>NPHV RNA</th>
<th>Antibodies</th>
</tr>
</thead>
<tbody>
<tr>
<td>Horses</td>
<td>PCR +</td>
<td>Ab+</td>
</tr>
<tr>
<td>Hepatopathy</td>
<td>0/111</td>
<td>53/111</td>
</tr>
<tr>
<td>No hepatopathy</td>
<td>3/217</td>
<td>89/217</td>
</tr>
<tr>
<td>Donkeys</td>
<td>0/100</td>
<td>0/100</td>
</tr>
<tr>
<td>Humans</td>
<td>0/362</td>
<td>0/362</td>
</tr>
<tr>
<td>Non-human primates</td>
<td>0/164</td>
<td>0/164</td>
</tr>
<tr>
<td>Dogs</td>
<td>0/113</td>
<td>1/113</td>
</tr>
<tr>
<td>Cats</td>
<td>0/131</td>
<td>0/131</td>
</tr>
</tbody>
</table>

*TABLE 6.2: NPHV RNA and antibodies detected by PCR and ELISA respectively.*
FIGURE 6.5: Demonstrated seroreactivity of serum tested in non-expressing control wells versus (A) NPHV-helicase, (B) NPHV-core and (C) EPgV-helicase antigen coated wells. Figure indicates that both ELISA positive and ELISA negative equine serum samples did not produce a higher OD in control wells coated with non-recombinant E. coli soluble fractions compared to the OD observed in wells coated with NS3 or core antigens from NPHV and EPgV. Seroreactivity to the control wells is denoted in red, while seroreactivity to the antigen-coated wells is denoted in blue. The established ELISA cut-offs are included for reference.

All samples were screened for both anti-NS3 and anti-core IgG antibodies. Samples were considered positive if above the cut-off in both ELISAs; those that were solely NS3 antibody positive or core antibody positive were classified as indeterminate (Table 6.3). A plot of the serology results for the NPHV NS3 OD against NPHV Core OD was used to validate the ELISAs (Figure 6.6) and demonstrated good concordance between reactivity to NS3 and core antigens (R2 = 0.5285) (Figure 6.7). Overall, 142 from 327 horse samples tested in both assays were confirmed positive (43.3%) and 19 (5.5%) were reactive in only one assay. In view of their indeterminate status, the latter 19 samples were excluded from subsequent analyses of disease status. All samples from horses previously identified as viraemic (Lyons et al. 2012), were reactive with both antigens by ELISA (Figure 6.6, Table 6.4).

TABLE 6.3: Seropositive horses determined by detection above cut-off of antibodies to both NS3 and Core NPHV proteins by ELISA. Samples that were seropositive in either NS3 or core ELISA were considered indeterminate (18 in total), samples above the threshold in both assays were considered antibody positive (142 horses).

<table>
<thead>
<tr>
<th></th>
<th>NPHV</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>NS3 Ab+</td>
</tr>
<tr>
<td>Core Ab+</td>
<td>142</td>
</tr>
<tr>
<td>Core Ab-</td>
<td>3</td>
</tr>
</tbody>
</table>

TABLE 6.4: Relative numbers of horses testing antibody positive to NS3 and Core ELISAs and/or RNA positive for NPHV. All RNA positive horses were antibody positive by both NS3 and Core ELISAs.

<table>
<thead>
<tr>
<th></th>
<th>NPHV</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>PCR+</td>
</tr>
<tr>
<td>Ab+</td>
<td>3</td>
</tr>
<tr>
<td>Ab-</td>
<td>0</td>
</tr>
</tbody>
</table>
FIGURE 6.6: Anti-NPHV helicase and core IgG detection across range of species. RNA positive cases are marked in red for horses, with repeat RNA positives in pink. RNA positive dog is marked with green triangle throughout. Samples labelled 1-3 refer to previously reported RNA positive horses while sample 4 refers to seropositive canine in regular contact with viraemic horse 3.
Seroprevalence of NPHV and EPgV in domestic horses

**FIGURE 6.7:** Assay validation of NPHV ELISA and correlation between samples identified as NS3 and/or Core antibody positive. Samples in red indicate previously reported viraemic horses, pink joined by black line indicates declining seroreactivity of a resampled viraemic over 1 year period horse (horse 3, Table 6.5), green indicates seropositive canine sample from the same farm.

All samples from all other mammalian species were unreactive in both NS3 and core ELISAs with the exception of a single sample taken from a dog that had come into regular contact with the previously reported NPHV viraemic horse (Table 6.5; ODs: NS3: 1.00, core: 0.63) (Lyons et al. 2012) (Chapter 5). Human serum samples from HCV positive patients did not react with the NPHV-NS3 or core ELISAs indicating no cross-reactivity between the NPHV antigens and HCV antibodies (Figure 6.6, 6.8).
FIGURE 6.8 Anti-NPHV helicase and core IgG detection in HCV positive and HCV negative human serum. Samples in red indicate absorbance of HCV positive samples in both the NS3 and core ELISAs. Reactivity of HCV negative serum is indicated in blue. No cross-reactivity between HCV antibodies and the NPHV proteins was detected.

Serial samples collected from a persistently viraemic horse showed declining antibody levels over a one year period (Figure 6.6, 6.7; Table 6.5). Over the initial 6 month period viral loads were elevated but declining as previously reported (Lyons...
et al. 2012). At the time of final sampling, RNA was undetectable indicating viral clearance but the horse remained seropositive, similarly all 5 in contact horses were non-viraemic but seropositive (Table 6.5). All 3 NPHV viraemic seropositive horses were from the control horse group (n=111), being clinically healthy and having no previous indication of hepatopathy. However, subsequent testing revealed that one had elevated serum gamma glutamyl transferase (59 U/L; upper reference level 42U/L) (Lyons et al., 2012a). The proportions of NPHV seropositive horses in the hepatopathy (53/111) and control (89/217) horse groups, 47.75% and 41% respectively, were not significantly different (Fisher’s exact test, p=0.29; Table 6.2).

**TABLE 6.5: NPHV NS3 and core antibodies detected in five horses and one dog cohabiting with an RNA positive horse.** Five horses that shared stables with horse EF369/11 (Table 6.5) (reported as actively viraemic between December 2011 and March 2012) were all antibody to NPHV, in addition to a single dog living on the farm and in daily contact with all horses.

<table>
<thead>
<tr>
<th>Host</th>
<th>Symbol (Figure 6.6)</th>
<th>Sample</th>
<th>Sample date</th>
<th>Viral load</th>
<th>NS3 NPHV OD</th>
<th>Core NPHV OD</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>EF_317/98</td>
<td>1998</td>
<td>1.3 x 10^5</td>
<td>0.65</td>
<td>0.75</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>EF_330/97</td>
<td>1997</td>
<td>4.4 x 10^5</td>
<td>0.55</td>
<td>0.66</td>
</tr>
<tr>
<td>3a</td>
<td>1</td>
<td>EF_369/11</td>
<td>Dec 2011</td>
<td>4.8 x 10^7</td>
<td>0.90</td>
<td>0.69</td>
</tr>
<tr>
<td>3b</td>
<td>1</td>
<td>EF_374/12</td>
<td>Mar 2012</td>
<td>2.1 x 10^5</td>
<td>0.67</td>
<td>0.68</td>
</tr>
<tr>
<td>3c</td>
<td>1</td>
<td>EF_523/12</td>
<td>Mar 2012</td>
<td>7.1 x 10^4</td>
<td>0.51</td>
<td>0.62</td>
</tr>
<tr>
<td>3d</td>
<td>1</td>
<td>EF_725/13</td>
<td>Dec 2012</td>
<td>Negative</td>
<td>0.50</td>
<td>0.41</td>
</tr>
<tr>
<td>3e</td>
<td>1</td>
<td>EF_726/13</td>
<td>Apr 2013</td>
<td>Negative</td>
<td>0.37</td>
<td>0.38</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>EF_600/13</td>
<td>July 2013</td>
<td>Negative</td>
<td>1.00</td>
<td>0.63</td>
</tr>
<tr>
<td>5</td>
<td>3</td>
<td>EF_520/12</td>
<td>Mar 2012</td>
<td>Negative</td>
<td>0.61</td>
<td>0.50</td>
</tr>
<tr>
<td>6</td>
<td>3</td>
<td>EF_519/12</td>
<td>Mar 2012</td>
<td>Negative</td>
<td>0.76</td>
<td>0.70</td>
</tr>
<tr>
<td>7</td>
<td>3</td>
<td>EF_517/12</td>
<td>Mar 2012</td>
<td>Negative</td>
<td>0.79</td>
<td>0.66</td>
</tr>
<tr>
<td>8</td>
<td>3</td>
<td>EF_521/12</td>
<td>Mar 2012</td>
<td>Negative</td>
<td>0.77</td>
<td>0.59</td>
</tr>
<tr>
<td>9</td>
<td>3</td>
<td>EF_518/12</td>
<td>Mar 2012</td>
<td>Negative</td>
<td>0.70</td>
<td>0.69</td>
</tr>
</tbody>
</table>

1RNA copies /ml determined by qRT-PCR
2Dog and 3horses that are in regular contact with Horse 3.
3Values corrected to two decimals places.
6.3.2 Preliminary estimation of seroprevalence of EPgV

An assay to detect antibodies to EPgV was developed by expression of the NS3 region in *E. coli* (pET-28b) that was homologous to that used for NPHV antibody screening. Recombinant protein was used in an indirect ELISA format using the same blocking and control antigens used for NPHV screening to minimise assay non-specificity. Due to the lack of a comparable core region for EPgV, western blots were used to confirm the presence of antibody in samples reactive to the EPgV-NS3 ELISA. Serum samples from 328 horses and 100 donkeys were initially tested (Table 6.6). Using an OD threshold of the mean serological reactivity (0.130) + 3 SDs of unreactive samples (0.262), 63% of samples from horses but none of the 100 donkey samples were identified as reactive by ELISA (Figure 6.9).

### TABLE 6.6: Detection of TDAV and EPgV RNA and antibodies by virus specific PCR and EPgV NS3 ELISA respectively.

TDAV RNA was not detected in horses or donkeys in the UK. EPgV RNA and antibodies were detected only in horses. Segregation of equine samples into hepatic and non-hepatic cases provides no evidence for a correlation between EPgV infection past or present and hepatitis.

<table>
<thead>
<tr>
<th>Species</th>
<th>EPgV</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>PCR+</td>
</tr>
<tr>
<td><strong>Horses:</strong></td>
<td></td>
</tr>
<tr>
<td>Hepatopathy</td>
<td>7/111</td>
</tr>
<tr>
<td>No hepatopathy</td>
<td>5/217</td>
</tr>
<tr>
<td><strong>Donkeys</strong></td>
<td>0/100</td>
</tr>
<tr>
<td><strong>Humans</strong></td>
<td></td>
</tr>
<tr>
<td>Non-human primates</td>
<td></td>
</tr>
<tr>
<td>Dogs</td>
<td></td>
</tr>
<tr>
<td>Cats</td>
<td></td>
</tr>
</tbody>
</table>
FIGURE 6.9: Anti-EPgV helicase IgG detected in horses by ELISA. All RNA positive cases previously identified (Chapter 5) are marked in red. Not all viraemic horses were seropositive by ELISA. RNA positive horses that were Western blot negative are denoted by (−) symbol in the figure above.
Screen-positive samples \((n=218)\) were tested for anti-NS3 antibodies by Western blot, of which 88\% were confirmed and considered positive (192/218 horses confirmed EPgV antibody positive). For the purposes of the subsequent analysis, NS3 ELISA-reactive samples that were negative in the supplementary WB assay were considered as indeterminate and not included in analyses of disease associations (Table 6.7).

**TABLE 6.7: Relative numbers of EPgV seropositive and seronegative horses testing antibody positive/negative by western blot.** Horses seropositive for EPgV antibodies by ELISA were confirmed by western blot. Samples western blot positive and negative were further segregated into horses presenting with or without biological indicators of hepatitis.

<table>
<thead>
<tr>
<th></th>
<th>EPgV ELISA Positive ((n=218))</th>
<th>EPgV ELISA Negative ((n=110))</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Hepatic</td>
<td>Non-Hepatic</td>
</tr>
<tr>
<td>Western Blot +</td>
<td>62</td>
<td>130</td>
</tr>
<tr>
<td>Western Blot -</td>
<td>12</td>
<td>14</td>
</tr>
</tbody>
</table>

To determine frequencies of active EPgV infection and to investigate the relationship between viraemia and presence of EPgV-specific antibody, all samples were screened by PCR using EPgV-specific primers targeting the NS3 region as outlined in Chapter 5. Viraemia was detected in 12/328 horses and ELISA screening of EPgV RNA positive samples detected antibodies in 11/12 horses (Table 6.8, 6.9). Secondary testing of these samples by western blot confirmed the presence of anti-EPgV NS3 antibodies in 10/12 horses (Table 6.9). Screening of serial samples from horses 1-5 detected antibodies by ELISA and western blot throughout the 4month sampling period during which viral RNA was detectable with viral loads ranging between \(1.16 \times 10^6\) and \(1.98 \times 10^9\) RNA copies/ml (Table 6.9). Proportions of EPgV
seropositive horses in the hepatopathy (62/111) and control (130/217) groups were not significantly different (Fischer’s exact test, p=0.55; Table 1).

**TABLE 6.8: The relative number of horses testing antibody positive and/or RNA positive for *EPgV*.** Eleven out of 12 *EPgV* RNA positive horse samples were seropositive when tested for antibodies by ELISA. Antibodies were not detected in RNA positive horse 8 despite high viral loads present.

<table>
<thead>
<tr>
<th>EPgV</th>
<th>PCR+</th>
<th>PCR -</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ab +</td>
<td>11</td>
<td>205</td>
</tr>
<tr>
<td>Ab -</td>
<td>1</td>
<td>119</td>
</tr>
</tbody>
</table>

Five out of 10 of the confirmed *EPgV* seropositive and viraemic horses were from the control horse group, being clinically healthy with no prior indication of hepatopathy. However subsequent testing of the viraemic control horses revealed that 2/7 had serum biochemical evidence of mild hepatopathy. Thus overall 7/10 *EPgV* seropositive viraemic horses had serum biochemical evidence of hepatopathy, namely elevation in GGT (6/10), GLDH (2/10) and/or bile acids (n=1)(Table 6.9), but all 6 for which there was available clinical information had no clinical signs of hepatopathy or other systemic disease. Repeat sampling of 5 of the initially viraemic horses demonstrated persistent infections were maintained over 4 month period although with declining viral loads and variable antibody levels (Table 6.9). Two of the 5 horses displayed overall elevations in NS3 antibody levels while the remaining 3 horses displayed mildly reduced seroreactivity with declining viral load over the 4 months measured. There was little evidence for hepatopathy maintained over that period with liver enzyme levels largely within the normal range with the exception of mildly elevated GGT/GLDH/bile acids in two cases (Horse 2 and 3;Table 6.9). At the time of final sampling the levels of GLDH and bile acids in horse 2 normalised and the elevated GGT and GLDH also declined in horse 3 with GGT falling back within the reference range. Proportions of *EPgV* seropositive horses in the
hepatopathy (62/111) and control (130/217) groups were not significantly different (Fischer’s exact test, p=0.55; Table 6.6).

**TABLE 6.9: Detection of EPgV NS3 antibodies by ELISA and confirmed by western blot in RNA positive horses.** Serial samples for five horses sharing the same pasture were available to assess the altering viral loads and seroreactivity over the course of 4 months (Chapter 5). Sample 8 presented with high viral load but was not seropositive while sample 7 was indeterminate, as antibodies were not detected by western blot.

<table>
<thead>
<tr>
<th>Horse</th>
<th>Test Group</th>
<th>Viral Load (copies/ml)</th>
<th>EPgV Ab OD</th>
<th>Western Blot</th>
<th>GGT&lt;sup&gt;1&lt;/sup&gt; (&lt;42U/L)</th>
<th>GLDH&lt;sup&gt;1&lt;/sup&gt; (&lt;12U/L)</th>
<th>Bile Acids&lt;sup&gt;1&lt;/sup&gt; (0-12 µmol/l)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Control</td>
<td>1.46 x 10&lt;sup&gt;8&lt;/sup&gt;</td>
<td>0.44</td>
<td>Positive</td>
<td>0</td>
<td>4</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2.79 x 10&lt;sup&gt;7&lt;/sup&gt;</td>
<td>1.39</td>
<td>Positive</td>
<td>2</td>
<td>4</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2.34 x 10&lt;sup&gt;6&lt;/sup&gt;</td>
<td>0.88</td>
<td>Positive</td>
<td>9</td>
<td>7</td>
<td>4</td>
</tr>
<tr>
<td>2</td>
<td>Control</td>
<td>4.88 x 10&lt;sup&gt;6&lt;/sup&gt;</td>
<td>1.00</td>
<td>Positive</td>
<td>18</td>
<td>12</td>
<td><strong>15.4</strong></td>
</tr>
<tr>
<td></td>
<td></td>
<td>3.19 x 10&lt;sup&gt;6&lt;/sup&gt;</td>
<td>0.75</td>
<td>Positive</td>
<td>17</td>
<td>12.5</td>
<td><strong>16</strong></td>
</tr>
<tr>
<td></td>
<td></td>
<td>2.91 x 10&lt;sup&gt;6&lt;/sup&gt;</td>
<td>0.91</td>
<td>Positive</td>
<td>23</td>
<td>9</td>
<td>10</td>
</tr>
<tr>
<td>3</td>
<td>Control</td>
<td>1.98 x 10&lt;sup&gt;9&lt;/sup&gt;</td>
<td>0.97</td>
<td>Positive</td>
<td><strong>44</strong></td>
<td>23</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3.19 x 10&lt;sup&gt;8&lt;/sup&gt;</td>
<td>0.83</td>
<td>Positive</td>
<td>42</td>
<td>27</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2.79 x 10&lt;sup&gt;7&lt;/sup&gt;</td>
<td>0.81</td>
<td>Positive</td>
<td>40</td>
<td><strong>21</strong></td>
<td>10.6</td>
</tr>
<tr>
<td>4</td>
<td>Control</td>
<td>3.06 x 10&lt;sup&gt;8&lt;/sup&gt;</td>
<td>0.85</td>
<td>Positive</td>
<td>21</td>
<td>4</td>
<td>6.8</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.26 x 10&lt;sup&gt;6&lt;/sup&gt;</td>
<td>0.81</td>
<td>Positive</td>
<td>23</td>
<td>2</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.16 x 10&lt;sup&gt;6&lt;/sup&gt;</td>
<td>0.79</td>
<td>Positive</td>
<td>23</td>
<td>2</td>
<td>6</td>
</tr>
<tr>
<td>5</td>
<td>Control</td>
<td>1.59 x 10&lt;sup&gt;8&lt;/sup&gt;</td>
<td>0.45</td>
<td>Positive</td>
<td>34</td>
<td>9</td>
<td>10.2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2.25 x 10&lt;sup&gt;7&lt;/sup&gt;</td>
<td>0.89</td>
<td>Positive</td>
<td>30</td>
<td>10</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4.09 x 10&lt;sup&gt;5&lt;/sup&gt;</td>
<td>1.02</td>
<td>Positive</td>
<td>34</td>
<td>8</td>
<td>10</td>
</tr>
<tr>
<td>6</td>
<td>Hepatopathy</td>
<td>1.26 x 10&lt;sup&gt;9&lt;/sup&gt;</td>
<td>1.30</td>
<td>Positive</td>
<td><strong>63</strong></td>
<td>8</td>
<td>6</td>
</tr>
<tr>
<td>7</td>
<td>Hepatopathy</td>
<td>1.15 x 10&lt;sup&gt;9&lt;/sup&gt;</td>
<td>0.36</td>
<td>Negative&lt;sup&gt;2&lt;/sup&gt;</td>
<td>156</td>
<td>50</td>
<td><strong>28</strong></td>
</tr>
<tr>
<td>8</td>
<td>Hepatopathy</td>
<td>1.26 x 10&lt;sup&gt;9&lt;/sup&gt;</td>
<td>0.12</td>
<td>Negative&lt;sup&gt;2&lt;/sup&gt;</td>
<td>136</td>
<td>0</td>
<td>6</td>
</tr>
<tr>
<td>9</td>
<td>Hepatopathy</td>
<td>NA</td>
<td>0.49</td>
<td>Positive</td>
<td><strong>126</strong>&lt;sup&gt;2&lt;/sup&gt;</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>10</td>
<td>Hepatopathy</td>
<td>NA</td>
<td>0.55</td>
<td>Positive</td>
<td><strong>749</strong>&lt;sup&gt;2&lt;/sup&gt;</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>11</td>
<td>Hepatopathy</td>
<td>NA</td>
<td>0.45</td>
<td>Positive</td>
<td><strong>243</strong>&lt;sup&gt;2&lt;/sup&gt;</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>12</td>
<td>Hepatopathy</td>
<td>NA</td>
<td>0.42</td>
<td>Positive</td>
<td><strong>120</strong>&lt;sup&gt;2&lt;/sup&gt;</td>
<td>NA</td>
<td>NA</td>
</tr>
</tbody>
</table>

<sup>1</sup>Elevated liver enzyme levels are highlighted in bold.

<sup>2</sup>Elevated levels in these cases attributed to toxins in the pasture based on veterinary inspection.

<sup>3</sup>Indicates horses NS3 antibody by ELISA but western blot negative, concluded as indeterminate.

<sup>4</sup>Values corrected to two decimal places.
Combined analysis of both the NPHV and EPgV antibody assays confirmed that 42 horses were seropositive for both viruses (Table 6.10). Of these 42 EPgV-NPHV antibody positive cases, 11 were from the hepatopathy group and the remaining 31 from the control group (no symptoms of hepatopathy), indicating no statistical significance between antibody detection and NPHV/EPgV infection and the presence of symptoms of hepatopathy (Fishers exact test $p=0.2983$)

**Table 6.10:** The relative number of horses confirmed as antibody positive for both EPgV and NPHV. A total of 42 horses were confirmed as seropositive for EPgV and NPHV antibodies, with 36 horses in total negative for antibodies to either virus.

<table>
<thead>
<tr>
<th>NPH</th>
<th>Ab +</th>
<th>Ab -</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ab +</td>
<td>42</td>
<td>100</td>
</tr>
<tr>
<td>Ab -</td>
<td>150</td>
<td>36</td>
</tr>
</tbody>
</table>

### 6.4 Discussion

#### 6.4.1 Serology based methods for the detection of NPHV and EPgV infection across species

Detection of NPHV infection was initially determined by PCR screening and deep sequencing (Kapoor *et al.*, 2011, Lyons *et al.*, 2012a). As the detection of RNA in respiratory and faecal samples can be attributed to ingestion of contaminated materials a subsequent serology based method for the detection of CHV-NS3 antibodies was applied to detect past infections (Burbelo *et al.*, 2012b). Recombinant protein was expressed from the helicase domain of CHV NS3 and used as an antigen in a luciferase based assay; luciferase immunoprecipitation system (LIPS) to screen a range of species including horses, pigs, cows and dogs (Burbelo *et al.*, 2012b, Burbelo *et al.*, 2012a).
Seroprevalence of NPHV and EPgV in domestic horses

Both the NPHV and EPgV assays used the evolutionary conserved helicase protein as the target antigen given the level of genetic diversity associated with RNA viruses including HCV. The NS3 helicase region is also known to correspond to a highly immunogenic region of HCV (Burbelo et al., 2010). In the study reported here recombinant proteins was expressed from the evolutionary conserved helicase NS3 and core regions of NPHV and used in ELISAs to determine seroprevalence of NPHV in horses, donkeys, cats, dogs, NHPs and humans. By applying two separate ELISA assays to detect NPHV antibodies, only samples positive in both would be considered seropositive, and therefore allowing for more accurate determination of seroprevalence in the studied populations, although with the possibility of higher background than the LIPS assay. Therefore for the determination of specificity and sensitivity the cut-off value of the LIPS and ELISA assays was calculated by comparable means and serum highly seropositive to both ELISAs or to LIPS assay was included as internal controls to standardise the protocols. Furthermore HCV positive samples were tested on both ELISA assays to ensure no cross-reactivity with the NS3 and core region of HCV (Figure 6.6, Figure 6.8). The weak seropositive reported in a single cow by the LIPS method could not be confirmed by second assay; although the use of dual NS3 and core ELISAs used in our study reduced the possibility of assay non-specificity. A similarly designed LIPS assay was also applied to screen 100 canines for CHV-NS3 antibodies in a separate study, and failed to detect antibodies or viral RNA in all samples (Bexfield et al., 2014). Serological screening of 180 bats using an anti-HCV immunoblot and immunofluorescence assay (IFA) has also recently been applied to detect novel hepaciviruses in the species (Drexler et al., 2013a). Diluted serum samples available from bats were tested against HCV infected HuH7-cells or replicon JFH1-transfected cells. Where sufficient samples were available IFA positive samples were tested using a commercial HCV immunoblot kit employing NS3, helicase, NS5, NS4 and core proteins. Sufficient sample volume was only available for 3/13 bat samples that
tested IFA seropositive and these were confirmed by the commercial immunoblot from Microgen.

In contrast to HCV where antibodies are detectable to several viral proteins during viraemia (Baumert et al., 2000), HPgV antibodies are not normally detected until clearance of viraemia, but several studies have reported the detection of antibodies to HPgV E2 and NS4 and NS5 (Fernandez-Vidal et al., 2007, Gomara et al., 2010, Pilot-Matias et al., 1996a, Schwarze-Zander et al., 2006, Tan et al., 1999, Van der Bij et al., 2005, Xiang et al., 1998). Most people infected with HPgV develop conformation-dependent antibodies to the E2 envelope glycoprotein, therefore serological screening employ E2 protein as a marker to detect past exposure and infection (Barnes et al., 2007, Nakatsuji et al., 1992, Tacke et al., 1997b, Tanaka et al., 1998a, Pilot-Matias et al., 1996a, McLinden et al., 2006, Gutierrez et al., 1997). The second assay in this study was designed to detect the presence of EPgV NS3 helicase antibodies, a novel approach for this virus, and with the lack of a comparable core region for EPgV all screen-positive samples were confirmed by western blot, again offering a secondary approach for confirming past exposure to the virus. NS3 helicase was chosen to enable comparison to be drawn with the NPHV assay and also to increase the possibility of detecting EPgV antibodies across species raised against an evolutionary conserved protein of the Pegivirus genus.

6.4.2 Serological evidence to suggest NPHV is endemic in the equine population

Determination of the seroprevalence of NPHV was carried out using dual ELISA assay based on the core and NS3, both regions of documented antigenicity in HCV. Additionally NS3 helicase domain is a highly conserved viral protein, allowing for the increased possibility of detecting more genetically diverse variants of NPHV. A seroprevalence of 43.3% was detected in horses, which is greater than the 35% previously reported in the United States, (Burbelo et al., 2012b) but not significantly
Seroprevalence of NPHV and EPgV in domestic horses

so (P=0.1380 Fishers exact test). These findings indicate NPHV is endemic and widely circulates in horse populations both in the UK/France and US. In regions of endemic HCV infection like Egypt and Pakistan seroprevalence rates of ≈4% and 15% respectively are observed in humans (Qureshi et al., 2010, Miller and Abu-Raddad, 2010, Averhoff et al., 2012), significantly below that observed for NPHV. This suggests possible differences in transmission efficiency, pathogenesis, tissue tropism and epidemiology or host factors between these two hepacivirus species. NPHV seroprevalence is also below initial reports of a 10.6 % seroprevalence in bats using an anti-HCV IFA and immunoblot, these findings remains to be confirmed using BHV specific assay (Drexler et al., 2013a). The level of NPHV seroprevalence observed in horses is closer to the levels observed amongst IDUs, and other risk groups. Studies have estimated 46% seroprevalence in IDUs in India (Basu et al., 2013), and a study of risk groups in Puerto Rico detected HCV seroprevalence rates of 39.2% among lifetime heroin users, 30.4% among individuals with past HBV infection, and 39.6% in lifetime cocaine users (Perez et al., 2005). Separation and analysis of seropositive equine samples based on presence or absence of hepatic symptoms provides no indication of increased levels of past exposure in animals presenting with elevated liver enzymes when compared to healthy animals (Fishers exact test p=0.1951).

All other species tested (Table 6.2) were seronegative with the exception of a single dog (Table 6.5, 3) that lives on a farm with a previously reported NPHV viraemic horse (Lyons et al. 2012) (Figure 6.6, Table 6.5). Identification of the first seropositive dog in the UK is the first significant indication that NPHV may be transferred to other species. This finding is significant and concordant with the initial report of CHV in the US, identified in association with an outbreak of canine respiratory infections in separate geographically isolated dog kennels (Kapoor et al., 2011). It is possible therefore that the initial CHV infections detected were the consequence of NPHV cross-species transmission to dogs. This is further supported
by subsequent studies failure to detect CHV RNA in UK canines (Lyons et al., 2012a, Bexfield et al., 2014).

Clearance of detectable levels of viral RNA after 6 months was observed in serial samples taken over 16 month period from a previously reported actively viraemic horse (EF369/11- Figure 6.6, 6.7), although antibodies remained detectable throughout the screening period and 6 months post clearance of RNA (Table 6.5). This is consistent with the regular detection of HCV antibodies to several viral proteins throughout the course of HCV viraemia; subsequently HCV antibodies decline over time and individuals may be seronegative (Takaki et al., 2000). Antibodies were also detectable in all stable mates of the viraemic horses further suggesting NPHV is endemic in the UK horses population, although with no presenting symptoms of hepatic disease or any noticeable impact upon horse performance, implying significant differences in the course of HCV and NPHV infections and their subsequent impact on host health.

6.4.3 Evidence for high levels of exposure to and clearance of EPgV

Based on the recent detection and characterisation of pegivirus infections in rodents, bats and horses (Simons et al., 1995a, Simons et al., 1995b, Epstein et al., 2010, Stapleton et al., 2011, Kapoor et al., 2013a, Kapoor et al., 2013b, Chandriani et al., 2013, Quan et al., 2013) a second EPgV specific assay was applied to determine seroprevalence in the UK. Anti-EPgV NS3 antibodies were detected in 58.5% of horses (192/328), 55.9% (62/111) in horses with symptoms of hepatitis and 60% (132/217) healthy animals (Figure 6.9; Table 6.8).

This exposure frequency is substantially higher that the seroprevalence of HPgV in humans, globally estimated at between 2-13% in healthy blood donors (Blair et al., 1998, Gutierrez et al., 1997, Pilot-Matias et al., 1996a, Tacke et al., 1997a). Higher rates have been described in groups with frequent sexual contact, such as up to 46% of HIV infected homosexual men (Williams et al., 2004, Pavlova et al., 1999) and
18% among prostitutes (Scallan et al., 1998b). To date no studies have detailed the seroprevalence in bats or rodents. A study on HPgV exposure in the US found that most individuals could be categorised as RNA positive/ELISA negative or ELISA positive/RNA negative and therefore calculated exposure as the sum of HPgV RNA positive and HPgV E2 antibody positive. Using this criteria exposure rates as high as 89.2% were observed in IDUs and as low as 5.5 in volunteer blood donors (Gutierrez et al., 1997).

Further studies on the seroprevalence of EPgV globally will therefore provide more information on viral transmission routes that may give rise to the levels of infection detected in this study. Given the careful veterinary practices observed in the modern day and the value attributed to horses both competing at international level or those kept as domestic pets it is difficult to attribute the seroprevalence observed to the transmission associated with IDUs. The recent detection of a second genetically distinct pegivirus in horses termed TDAV, associated with the receipt of a botulinum antitoxin suggests the transmission of these viruses may be associated with plasma and blood transfusions (Chandriani et al., 2013). In such instances the commercial screening of plasma administered to horses for the presence of novel viruses such as TDAV or EPgV may be required but is not yet available or not yet deemed a requirement due to a lack of understanding into tissue tropism and disease associations (Aleman et al., 2005, Thomsett, 1971, Guglick et al., Panciera). The novel assays described within this chapter provide initial potential methodologies for the successful screening of equine blood products, and indeed equine populations globally in order to better understand the transmission routes and disease association of these novel equine specific viruses.
6.4.4 Assessment of the correlation between NPHV and EPgV antibody detection and symptoms of hepatic disease

Based on the segregation of equine samples in this study into animals presenting with or without evidence of hepatitis, of the 142 horses confirmed seropositive for NPHV, 48.6% (54/111) of hepatic horses were antibody positive to NPHV and 40.6% (88/217) of non-hepatic (Fishers exact test p=0.1951). Previous studies detected seroprevalence in horses of 35% (Burbelo et al., 2012b) but no details are available to enable a breakdown of this 35% into hepatitis symptomatic and asymptomatic sub-groups. In the same study studies NPHV RNA was detected in 20% (8/36) of seropositive animals, although viraemia was only detected in 2% (3/142) of NPHV antibody positive horses in this study, a statistically significant difference (Fishers exact test p=0.0002). The detection of viraemia in 20% of antibody positive horses indicates potential persistent infection but both that study and data presented here provides no indication of a correlation between antibody detection and liver disease, a significant contrast to all we know of HCV infections and the definitive liver tropism and hepatic disease association.

All RNA positive horses were antibody positive (Table 6.5) consistent with HCV infections (Takaki et al., 2000), although analysis of liver enzymes in all viraemic animals fails to indicate hepatic insufficiencies (Chapter 4). Furthermore based on discussion with the horse owner of the previously reported viraemic Horse 3 (Table 6.5, 6.4) there was no evident shift in the horse’s mood, behaviour, or performance proceeding, during or after the 16 month sampling period. Similarly comparison of the confirmed rates of seroprevalence for EPgV amongst hepatic 55.9% (62/111) and non-hepatic 59.9% (132/217) animals indicates no correlation between the rates of antibody detection and hepatic disease (Fishers exact test p=0.5538). As this is the first study into the seroprevalence of EPgV in the domestic horse population it is an early indication that EPgV infection is endemic in the population.
Antibodies were detected in 10/12 actively viraemic horses, and of those horses confirmed as EPgV seropositive 5.2% (194/328) were RNA positive. Detection of antibodies and high viral loads over 4 month period in Horses 1-5 (Table 6.9) in parallel suggests the presence of persistent infections. Persistence of infection is characteristic of other pegiviruses, including SPgV (Simons et al., 1995b) and also EPgV in horses in the US where infection was documented over a period of at least 3.5 years in two horses (Kapoor et al., 2013a). Similarly, while immune competent individuals typically clear HPgV infections, this occurs over relatively prolonged timeframes, years after primary infection (Tanaka et al., 1998a, Berg et al., 1999). Over the 4-month period within which horse samples were collected (Table 4), antibody levels were generally maintained despite a general declining trend in viral load. These findings contrast with reports of absent anti-E2 antibody in human actively infected with HPgV and seroconversion on clearance of viraemia (Gomara et al., 2010, Pilot-Matias et al., 1996b, Schwarze-Zander et al., 2006, Tan et al., 1999, Van der Bij et al., 2005). Studies on persistence of HPgV infection have detected RNA over 3 years post infection (Gutierrez et al., 1997) but comparable studies on EPgV infection are required before conclusions on course of infection can be drawn. Furthermore based on the analysis of horses presenting as antibody positive to both EPgV and NPHV (n=42) (Table 6.10), just over a quarter of these cases (11/42, 26%) were associated with the hepatopathy group (n=111). The remaining 31/42 cases (74%) were from the control group (n=217). Based on these preliminary findings there is no indication of a correlation between prior infection with both viruses and an increased risk of hepatopathy (Fishers exact test p=0.2983).

Of the antibody positive/RNA positive horses liver enzymes levels analysed were not found to be elevated to levels indicative of liver insufficiencies (Table 6.9). Persistent infections have been documented for another distinct equine pegivirus TDAV, where RNA has been detected in serum 1-year post infection (Chandriani et al., 2013). TDAV has been associated with the occurrence of acute serum hepatitis in
Seroprevalence of NPHV and EPgV in domestic horses

horses and future determination of the seroprevalence of TDAV in horses presenting with hepatic disease will provide another indication of the presence of absence of a roll for pegiviruses in equine hepatitis (Thomsett, 1971, Aleman et al., 2005, Panciera, 1969, Guglick et al., 1995).

6.4.5 Conclusions

Findings presented here suggest NPHV and EPgV are endemic in the UK equine population with no determined disease association or correlation with liver enzyme levels. Observation of liver enzymes and viral loads provided no definitive indications that the detection of NPHV or EPgV antibodies was correlated to hepatic disease in horses. To date TDAV is the only suggested viral link between acute serum hepatitis and pegiviruses. However the endemic nature of NPHV and EPgV in the domestic horse population as suggested by the data presented here warrants further investigation into the transmission routes, pathogenesis and tissue tropism of these viruses.

The finding of the first NPHV seropositive canine indicates that the original detection of CHV may have been a cross-species transmission event (Kapoor et al., 2011). While this study finds no direct correlation between NPHV or EPgV and equine hepatitis, it does provide possible tools to aid in the development of diagnostic kits to aid research on transmission, disease association and tissue tropism. The presence of hepacivirus and pegivirus species in dogs, horses, rodents and bats sheds light on the possible evolutionary history of HCV and HPgV, which consequentially raises the question of cross-species transmission and zoonosis. This study provides tools for the development of diagnostic assays for viraemia and antibody screening that will assist future research into their transmission, disease associations and tissue tropisms. The presence of hepacivirus and pegivirus species in dogs, horses, rodents and bats sheds light on the possible evolutionary history of HCV and HPgV, in which potential cross-species transmission and zoonotic origins
suddenly become more plausible. Continued PCR and serology based screening of other mammalian taxa across range geographical locations, including areas of endemic HCV infection is required to address these gaps in our understanding of the evolution of hepaciviruses and pegiviruses.

The recent detection of highly divergent hepaciviruses and pegiviruses in bats and rodents raises substantial questions relating to the ultimate origins HPgV and HCV and the role that these novel hosts (horses, rodents and bats) have played in their evolutionary history. It is essential that serological screening is carried out on these species in particular in regions associated with endemic HCV and HPgV infections and areas where human populations are exposed to or in contact infected bats, rodents, NHPs and other equine species. Through effective monitoring of the level of seroprevalence and active and past infections in mammalian species it will be possible to elucidate the origins, evolution, and dissemination of HCV and HPgV globally, develop effective health management and control strategies, and measure the determine the potential threat if any posed by the presence of hepacivirus and pegivirus reservoir in bats.
Chapter 7
Concluding remarks

7.1 Global importance of elucidating the origins of HBV and HCV

The global dissemination and ultimate origins of the major human hepatic pathogens HBV and HCV have been extensively studied, primarily as this information would impact the development of clinical treatments and predict the future evolution and spread these viruses. A major limitation of these investigations is the lack of available historical samples and evidence of infection; therefore the studies rely upon indirectly inferring origins and patterns of viral dissemination from epidemiological and phylogenetic analysis of currently circulating viruses among different species.

Infections with pathogenic HBV and HCV are globally distributed, and are major causative agents of chronic liver disease and hepatocellular carcinoma. Infections with HPgV are non-pathogenic but comparable with HBV and HCV are capable of establishing persistent and lifelong infections, with high levels of detectable virus replication. HBV, HCV and HPgV chronically infecting approximately 5%, 2% and 5% of the world’s population respectively. There is a significant global health burden associated with the health management of HBV and HCV infection, which equates to 350 million and 125 million people respectively (Perz et al., 2004, Simmonds, 2001a). One millions deaths per year are attributed to viral hepatitis, of which 78% are the result of HBV and HCV. In the UK alone it is estimated that 0.2% of the population is infected with HCV. However we have yet to see the full impact upon the health service of future liver associated complications related to these patients. If all those infected develop chronic liver disease the health burden would not be supportable by the current health services of the country. In order to improve future health management and strategies, and improve available treatments, it is vitally important that we better understand the origins of HCV.
Concluding Remarks

All three viruses show substantial heterogeneity and are classified into 7-8 genotypes. Tracing the origins of these viruses would allow for not only improved health treatment strategies but may also aid in the development of more effective global control and eradication measures. In order to determine the evolutionary history and examine past transmission of HCV, HBV and HPgV, the rate of sequence of different genotypes has been used to infer the estimated time of divergence. This method predicts a relatively recent emergence of these viruses into humans, HBV≈3000 years, HCV ≈500-2000 years and HPgV ≈200 years ago. These estimates however are incompatible with the epidemiological distribution of these viruses among the human population, and are further incompatible with the recent detection of species-specific variants of HBV and HPgV among NHPs, and bats, in addition to the detection of species-specific variants of HCV and HPgV among horses, and rodents, and New World monkey with respect to HCV. Recombination has also been found to occur between human and NHP variants of HBV, generating novel strains, and increasing the genetic diversity of the virus, and variants infecting 3 species of bats (RBHV/TBHBV/HBHBV) have been shown to be capable of successfully infecting and replicating in human hepatocytes.

The detection of these viruses among diverse mammalian species and the potential for recombination between variants and transmission between species significantly affects the long-term control and eradication of these viruses. Phylogenetic analysis places bats at a basal position to all equine, human and NHP variants suggesting that bats are potentially a major ancient natural reservoir for hepaciviruses and pegiviruses, potentially transmitted and maintained in regions of endemic infection through hunting, exposure to and consumption of wild bush meat. Global sampling and clinical analysis of infected animals at present suggests no definitive health impact to bats or horses infected with hepaciviruses or pegiviruses. However they provide new modes through which to model human infections and the potential to expand current treatment therapies through monitoring of infection in these novel...
hosts. The importance of these viruses to human health is evident in the rapid developments and detection within the past 3 years alone resulting in the characterisation of 9 new virus species across 2 virus families: the Hepadnaviridae and Flaviviridae, and 3 genera: Orthohepadnavirus (HBHBV/TBHBV/RBHBV), Flavivirus (CHV/NPHV/RHV/BHV/GHV) and Pegivirus (EPgV/RPgV/BPgV).

### 7.2 Novel findings of HBV recombination and species association among NHPs

Co-infection with two different genotypes of HBV in one individual or animals has been previously documented to result in the formation of recombinant variants of HBV. There are several stages during the viral life cycle at which recombination can occur including: RNA-RNA during the packaging of pgRNA or during the reverse transcription and the generation of linear double stranded genomes (Yang and Summers, 1995). Studies into HBV recombination have detected that the majority (90%) have occurred between genotypes B/C (Bollyky et al., 1996, Morozov et al., 2000, Sugauchi et al., 2002b, Mukaide et al., 1992) and A/D, (Bollyky et al., 1996, Bowyer and Sim, 2000, Morozov et al., 2000, Owiredu et al., 2001b) in addition to other recombinants A/E, A/G, C/D, C/F, C/G, C/U and B/C/U (Simmonds and Midgley, 2005, Yang et al., 2006). A single study has further documented the occurrence of recombination between the human genotype C and the chimpanzee variant AF498266 (Magiorkinis et al., 2005) and gibbon variants (Sa-Nguanmoo et al., 2009). Their occurrence demonstrates that despite their genetic divergence, human and non-human associated variants of HBV can share hosts in nature.

Studies documented in thesis (Chapter 3) focused on HBV variants circulating in NHPs in Cameroon, a region of endemic HBV infection and where NHPs and humans share environments. While no NHPs were found carrying human HBV genotypes, a unique recombination event was documented between a gorilla and
chimpanzee HBV variant. This implies that despite the species-specific clades observed for HBV NHP variants, that a gorilla can carry a chimpanzee variant, which in this particular instance was co-infected with a gorilla variant resulting in a recombinant. Furthermore this event was likely to have occurred in the wild as in captivity the two species are not cohoused. The additional observation through analysis of the *Pan troglodytes* sub-species of HBV positive samples that all HBV isolates grouped within the *P.t.elliotti* clade, irrespective of their sub-species host proposed that chimpanzee sub-species sharing overlapping habitats may carry either of the NHP variants. However as at the time of sampling all chimpanzees were co-housed, it is not possible to rule out if the sub-species transmission occurred in captivity or the wild. Further studies are required to determine if this occurs in the wild chimpanzee population, in particular is regions where habitats crossover, for example the Sanaga River.

As a result of these studies we proposed that HBV evolution might be the result of allopatric or geographical speciation and not virus/host co-evolution. This alternative hypothesis would explain the internal branching position of the gorilla clade within the chimpanzee derived HBV sequences and the inlier position of orangutan-derived sequences deep within the gibbon clade, and their close genetic relationship with the sympatric *Hylobates agilis* gibbon species (Sall et al., 2005, Noppornpanth et al., 2003) as these two species occupy proximal or overlapping habitats in Borneo, while HBV variants infecting gibbons from elsewhere in Asia group separately (Starkman et al., 2003, Warren et al., 1999). Based on the more recent detection of novel hepadnavirus variants in bats in Africa and Central America (Drexler et al., 2013b), and the implication that these species may be reservoirs for primate HBV warrants further investigation. In particular the future search for recombinant variants and cross-species transmission events should focus on geographical regions of Africa and other countries were the humans and NHPs may be in regular contact with bats, as a result of habitat, hunting and the consumption of bush meat.
7.3 Significance of detection and clinical characterisation of NPHV infection

Initially the main focus of the study was designed at screening NHPs for HCV homologs with the aim of tracing the origins of human infections to NHPs comparable to that observed for HIV-1 and HIV-2 (Gao et al., 1999, Feng et al., 1992). Fortuitously at this time came the identification of CHV, a canine homolog for HCV, isolated from respiratory samples of dogs with severe respiratory infection Kapoor (Kapoor et al., 2011) and the subsequent characterisation of an equine homolog for HCV, termed NPHV (Burbelo et al., 2012b).

As a consequence a revised study was designed to screen a range of mammals, including dogs, cats, mice, horses and NHPs for CHV, NPHV and HCV infections (Chapter 4). The study provided the first documented NPHV infections in horses outside the United States (Lyons et al., 2012a) and support the finding that horses are the natural hosts for NPHV. The finding of no CHV infections in dogs in UK was supported by subsequent studies, and provided evidence that the CHV detected may have been the result of cross-species transmission from horses to dogs. Phylogenetic analysis of the 3 novel UK isolates confirmed them as distinct from all available sequences NPHV. As studies rapidly refocused around the world on novel mammalian hosts for HCV homologs, divergent homologs were detected in rodents, bats and New World monkeys (Kapoor et al., 2013b, Quan et al., 2013, Lauck et al., 2013). Phylogenetic analysis based on the highly conserved NS3 and NS5B of all novel hepacivirus variants (Chapter 4) determined that NPHV was the most closely related to HCV, but as other studies had postulated, confirmed that bats may be ancient natural reservoirs of hepacivirus genetic diversity (Quan et al., 2013).

This study further allowed for the first initial studies into NPHV infection and its clinical impact up on the host. Through extensive monitoring of viral load, liver enzymes and overall behaviour of a single Scottish viraemic horse over 14month
period we were found that the viral load declined and within 6 months was finally undetectable with no adverse effects on the animal or any biological evidence of liver or other systemic disease. Such findings are in stark contrast to the vast majority of HCV infections that result in a chronic disease state in approximately 70-80% of cases. Our findings are more comparable to an acute self-limiting infection but the absence of any liver enzyme elevations is again unusual and points to certain host or viral characteristics that might render the infection non-pathogenic to the horse (Barrera et al., 1995).

Extensive future studies on horses suffering from hepatic disease or presenting with clinical evidence of disease for the presence of NPHV will more conclusively determine whether or not NPHV is hepatotropic and what role the horse immune response plays in disease progression. Furthermore the screening of horse, rodent and bat species in the UK as well further afield in regions of endemic HCV infection (sub-Saharan Africa) will further our understanding of the evolution and global dissemination of HCV.

7.4 Epidemiology and clinical significance of EPgV

Until infection with HCV and its genetically related GBV-B virus was, until recently, thought to be confined to humans and experimentally infected New World primates respectively. The detection of CHV and NPHV in dogs and horses respectively provided the first indication of a much wider host range for hepacivirus infection (Burbelo et al., 2012b, Lyons et al., 2012a, Kapoor et al., 2011) and as a consequence studies also looked to the other genus of the Flaviviridae family, the Pegivirus genus. Identification of an equine homologs for HPgV circulating in horses marked the beginning of studies into the detection and characterisation of divergent pegiviruses circulating in rodents and bats across the globe (Kapoor et al., 2013b, Kapoor et al., 2013a, Quan et al., 2013).
Findings presented within this thesis (Chapter 5) provide the first evidence of EPgV infections in horses in the UK and France with an as-yet undetermined disease association or correlation with liver enzyme levels. All EPgV sequence isolates reported here are distinct from those previously reported, in addition to TDAV; associated with acute serum hepatitis in horses. In contrast to TDAV findings in this study detected no significant impact on equine health by viral infection despite the high levels of viraemia. Active pegivirus infection was detected in 12/328 (3.8%) horses, within the range observed in preliminary studies of species-specific variants in rodents, bats and other horse surveys (3.9-9.5%) (Quan et al., 2013, Kapoor et al., 2013a, Kapoor et al., 2013b) and comparable with published data which estimates infection frequencies in the human population of ≈5% in the developed world with upwards of 20% viraemia recorded in some developing countries (Polgreen et al., 2003, Mohr and Stapleton, 2009, Pavlova et al., 1999). Failure to detect any active pegivirus infections in donkeys does not rule out the possibility that a species specific variant is circulating in the population, as has been observed for other members of this genus.

Examination of the clinical impact of EPgV infection relied upon examination of viral loads in serial samples and the monitoring of liver enzymes. As no liver biopsies or lymph fluid was available for testing the exact tissue tropism of the virus could not be conclusively determined. Liver enzymes were largely within the normal range with some exceptions, and where elevations were evident there was no biological evidence in the horse to suggest hepatic disease, although this cannot be ruled out. The US study reported the detection however of EPgV in liver tissue, suggesting potential hepatotropism (Kapoor et al., 2013a). Indeed when data for hepatic and healthy controls in this study was combined with that observed in the US study (Kapoor et al., 2013b), there was statistically significant evidence for an association between hepatopathy and EPgV viraemia; with combined infection frequencies of 8.1% (10/123) in horses with hepatopathy compared to 3.2% (9/279) in control horses (Fisher’s exact test; p=0.042). This correlation provides very
preliminary data suggesting that EPgV may be a cause or contributory factor in equine hepatopathy. This is contrary to what we observe for other member of the Pegivirus genus, reclassified as such partly as a consequence of their non-hepatic disease associations (Stapleton et al., 2011). The recent detection of a novel and highly divergent hepacivirus in an OWM highlights the possibility that divergent and as yet undetected hepaci-and pegi-like viruses are present and circulating in the wild amongst NHPs and other species (Lauck et al., 2013). Furthermore our observation of high viral loads in five horses over a 5-month period suggests the ability of EPgV to establish persistent infections in the host without clinical symptoms. The new discoveries of a much wider and diverse range of pegiviruses reported in horses, bats and rodents across the globe fundamentally revises our knowledge of viral diversity and host ranges of viruses in this genus, their epidemiology and pathogenesis. (Quan et al., 2013, Kapoor et al., 2013b, Kapoor et al., 2013a). Continued PCR of other mammalian taxa across a range of geographical locations, including areas of endemic HPgV infection is required to address these gaps in our understanding of pegivirus evolution.

7.5 Novel serological assays determine endemic levels of NPHV and EPgV among horses

Following the detection of active NPHV and EPgV infections in horses (Chapter 4 and 5) we sought to determine the seroprevalence of both viruses in the same population. Using RNA positive samples we designed novel ELISA assays based on NPHV NS3 helicase and Core and EPgV NS3 helicase regions. Through the examination of the level of exposure in the horse population we aimed to expand our understanding of the host range, transmission routes, and disease progression of both viruses (Chapter 6).
Concluding Remarks

We firstly determined an NPHV seroprevalence of 43.3% in horses, slightly greater than the 35% reported in the United States, (Burbelo et al., 2012b) but not significantly so (P=0.1380 Fishers exact test). These findings provide the first initial evidence that NPHV is endemic and widely circulates in horse populations both in the UK/France and US at frequencies far greater than those observed even in regions of endemic HCV infection like Egypt and Pakistan (seroprevalence rates of ≈4% and 15% respectively) (Qureshi et al., 2010, Miller and Abu-Raddad, 2010, Averhoff et al., 2012). This suggests possible differences in transmission efficiency, pathogenesis, tissue tropism and epidemiology or host factors between these two hepacivirus species. The level of NPHV seroprevalence observed in horses is closer to that observed amongst risk groups like IDUs in which studies have detected exposure rates of up to 46% (Basu et al., 2013, Perez et al., 2005).

A surprising finding of the NPHV study was the detection of the first NPHV seropositive canine that has for many years lived on the same farm as the previously reported viraemic horse (Lyons et al., 2012a). This provides the first supporting evidence of the initial detection of CHV in canines in the US and highlights how potentially the specific transmission routes of this virus make this jump between species a significantly rare event. The subsequent detection of divergent hepaciviruses in bats, rodents and New World monkeys reinforces the fact that these viruses have a far greater diversity and host range than previously thought and that a lot remains to be determined. The newly emerging hypothesis is that bats may now in fact be ancient reservoirs for hepacivirus diversity and the ultimate origin of HCV in humans. Surveys into the seroprevalence of these novel viruses in the host population will improve our understanding of their evolution, host range, and pathogenesis.

The novel EPgV assay detected anti-EPgV NS3 antibodies in 58.5% of horses (192/328), 55.9% (62/111) in horses with symptoms of hepatitis and 60% (132/217) in healthy animals. This is substantially higher that the seroprevalence of HPgV in
the human population, globally estimated at between 2-13% in healthy blood donors (Blair et al., 1998, Gutierrez et al., 1997, Pilot-Matias et al., 1996a, Tacke et al., 1997a). However, much higher rates have been described in groups with frequent sexual contact, such as up to 46% of HIV infected homosexual men (Williams et al., 2004, Pavlova et al., 1999) and 18% among prostitutes (Scallan et al., 1998b). Again as with NPHV the method of transmission or host factors might play a significant role in the course of EPgV infection. The detection of antibodies during viraemia is not generally associated with pegivirus infections although some studies report detecting anti-HPgV peptide antibodies (Fernandez-Vidal et al., 2007, Gomara et al., 2010, Pilot-Matias et al., 1996a, Schwarze-Zander et al., 2006, Tan et al., 1999, Vander Bij et al., 2005, Xiang et al., 1998). Based on the EPgV seroprevalence detected in diseased versus control horses there was no indication of a correlation between the increased antibodies detection and presence of hepatic disease.

This study provides effective tools for the development of diagnostic assays for viraemia and antibody screening that will assist future research into transmission, disease associations and tissue tropism of NPHV and EPgV.

### 7.6 Concluding remarks

In conclusion, the research undertaken during this PhD has sought to expand our understanding of the host range, epidemiology, origins and evolution of the human hepatic viruses, HBV and HCV and the genetically related HPgV. We have proposed an alternate hypothesis of allopatric speciation for the evolution of HBV (Chapter 3) and identified a novel gorilla/chimpanzee recombinant of wild origins from a region of endemic HBV infection, Cameroon. The successful isolation of NPHV, an equine homolog of HCV in the UK, allowed for the very first preliminary investigations into the course of NPHV infection, its clinical impact upon the host and as a consequence concluded that at present there was no evidence to suggest that NPHV infection was
hepatotropic, in stark contrast to HCV (Chapter 4). Our studies then looked at the Pegivirus genus and the equine homolog EPgV. Combining our data relating to active infections with that observed in the US, we caution to suggest that EPgV might be one of several factors associated with hepatic disease in horses. The extensive NPHV and EPgV seroprevalence study we carried out confirmed the observed species-specific association of both viruses and that NPHV and EPgV were endemic in the horse population at rates exceeding those observed for HCV and HPgV. The important detection of the first NPHV seropositive canine outside of the US study that first identified CHV highlights the rare but potential occurrence of cross-species transmission of NPHV to dogs, through as yet undetermined transmission routes.

Given how little was known about the origins, evolution and host range of hepaciviruses and pegiviruses at the initial stages of this PhD, the past three years has seen an unprecedented level of valuable and insightful research in each of these of these fields. The identification of novel hosts for these viruses in horses, rodents and bats has for the first time provided us with potential animal models for HCV and HPgV infection. Although at present we know little of the pathogenesis and tissue tropism of these viruses, future studies would ideal seek to examine these gaps in our understanding. Furthermore, the availability of novel assays and sequence data will allow for extensive screening of mammalian species around the world for novel viral pathogens homologous to HCV and HPgV. Ideally studies will focus on regions of endemic human infections, where species known to harbour homologs are in frequent contact with the human population, for example regions of central Africa where humans hunt and consume bat bushmeat and share their environment with NHPs increasing the potential for zoonotic transmission. Seroprevalence studies in areas such as sub-Saharan Africa and South East Asia will answer many of our questions relating to the transmission routes of these novel viruses. The availability of sequence data for a divergent New World monkey hepacivirus will reinforce the
importance of screening wild NHPs. In the long-term these studies combined with those of others over the past three years have the potential to shed extensive new light on the origins of some of the world’s major human pathogens and ultimately will impact upon measures taken to control, and/or eradicate these viruses and improve the availability of effective treatment regimes.
References


ANDRE, F. 2000. Hepatitis B epidemiology in Asia, the Middle East and Africa. Vaccine, 18 Suppl 1, S20-2.


References


FOGEDA, M., NAVAS, S., MARTIN, J., CASQUEIRO, M., RODRIGUEZ, E.,
AROCENA, C. & CARRENO, V. 1999. In vitro infection of human peripheral

FORGUES, M., DIFILIPPANTONIO, M. J., LINKE, S. P., RIED, T., NAGASHIMA, K.,
of Crm1 in hepatitis B virus X protein-induced aberrant centriole replication

FORGUES, M., MARROGI, A. J., SPILLARE, E. A., WU, C. G., YANG, Q., YOSHIDA,
M. & WANG, X. W. 2001. Interaction of the hepatitis B virus X protein with the

FOSTER, G. R. & THOMAS, H. C. 1993. Recent advances in the molecular biology of

of the Hepatitis-B Virus P-Gene by Ribosomal Scanning as an Alternative to

FOY, E., LI, K., WANG, C., SUMPTER, R., JR., IKEDA, M., LEMON, S. M. & GALE,
M., JR. 2003. Regulation of interferon regulatory factor-3 by the hepatitis C virus

FRANK, C., MOHAMED, M. K., STRICKLAND, G. T., LAVANCHY, D., ARTHUR, R.
R., MAGDER, L. S., EL KHOBY, T., ABDEL-WAHAB, Y., ALY OHN, E. S.,
ANWAR, W. & SALLAM, I. 2000. The role of parenteral antischistosomal therapy

FRETZ, C., JEANNEL, D., STUYVER, L., HERVE, V., LUNEL, F., BOUDIFA, A.,
MATHIOT, C., DE THE, G. & FOURNEL, J. J. 1995. HCV infection in a rural
population of the Central African Republic (CAR): evidence for three additional

FRICK, D. N., RYPMA, R. S., LAM, A. M. & GU, B. 2004. The nonstructural protein 3
protease/helicase requires an intact protease domain to unwind duplex RNA


FRIEMAN, J. S. & COSSART, Y. E. 1986. Natural duck hepatitis B virus infection in

GALE, M., JR., BLAKELY, C. M., KWIECISZEWSKI, B., TAN, S. L., DOSSETT, M.,
TANG, N. M., KORTH, M. J., POLYAK, S. J., GRETSCH, D. R. & KATZE, M. G.
1998. Control of PKR protein kinase by hepatitis C virus nonstructural 5A protein:


GERROS, T. J. 1996. Diseases of the gallbladder, cholangitis, cholangiohepatitis. , Mosby, St. Louis, USA.


GRETHE, S., GEMSA, F., MONAZAHIAN, M., BOHME, I., UY, A. & THOMSSSEN, R. 2000a. Molecular epidemiology of an outbreak of HCV in a hemodialysis unit:


273


LANFORD, R. E., CHAVEZ, D. & BRASKY, K. 1998. Isolation of a hepadnavirus from the woolly monkey, a New World primate.


References


LI, C., LU, L., WU, X., WANG, C., BENNETT, P., LU, T. & MURPHY, D. 2009. Complete genomic sequences for hepatitis C virus subtypes 4b, 4c, 4d, 4g, 4k, 4l, 4m, 4n, 4o, 4p, 4q, 4r and 4t. *J. Gen. Virol.*, 90, 1820-6.


References


References


References


289


References


SHIMAKAMI, T., HIJKIKA, T., Luo, H., MA, Y. Y., KANEKO, S., SHIMOTOHNO, K. & MURAKAMI, S. 2004. Effect of interaction between hepatitis C virus NS5A
References

and NS5B on hepatitis C virus RNA replication with the hepatitis C virus replicon. *J. Virol.*, 78, 2738-2748.


References


characteristics of hepatitis B virus genotype B having the recombination with genotype C. *Gastroenterology*, 124, 925-932.


References


C (hepatitis G virus) with viral clearance and protection from reinfection. *J. Infect. Dis.*, 177, 539-42.


TOKITA, H., OKAMOTO, H., IIZUKA, H., TSUDA, F., SHRESTHA, S., MIYAKAWA, Y. & MAYUMI, M. 1996. Hepatitis C virus variants from Jakarta, Indonesia classifiable into novel genotypes in the second (2e and 2f), tenth (10a) and eleventh (11a) genetic groups. *J. Gen. Virol.*, 77, 293-301.


Evolutionary history, cross-species transmission and host adaptation of human viruses and their primate homologues

Sinéad Lyons

Appendices

A thesis submitted for the degree of Doctor of Philosophy

The University of Edinburgh
### Appendix A – List of Primers

<table>
<thead>
<tr>
<th>Primer</th>
<th>Virus</th>
<th>Position</th>
<th>Sequence 5’→ 3’</th>
</tr>
</thead>
<tbody>
<tr>
<td>EQPGV_NS3OS</td>
<td>EPgV</td>
<td>OS</td>
<td>GGACGCAGACACGCCYCAAYCC</td>
</tr>
<tr>
<td>EQPGV_NS3IS</td>
<td>EPgV</td>
<td>IS</td>
<td>GGTGATTCGCCCTTCAGTGCGG</td>
</tr>
<tr>
<td>EQPGV_NS31AS</td>
<td>EPgV</td>
<td>IAS</td>
<td>ATTTATGCACGAGGGTCACCTTGCGC</td>
</tr>
<tr>
<td>EQPGV_NS3OAS</td>
<td>EPgV</td>
<td>OAS</td>
<td>GGTGCTGACACCGACAGTRGC</td>
</tr>
<tr>
<td>EQPGV_5UTROS</td>
<td>EPgV</td>
<td>OS</td>
<td>GGACGCAGAGCACGCCYCAAYCC</td>
</tr>
<tr>
<td>EQPGV_5UTRIS</td>
<td>EPgV</td>
<td>IS</td>
<td>GAGATACCCTGACTGCTGARGG</td>
</tr>
<tr>
<td>EQPGV_5UTRIAS</td>
<td>EPgV</td>
<td>IAS</td>
<td>AAGGAGGGATCTGGAGGATATC</td>
</tr>
<tr>
<td>EQPGV_5UTROAS</td>
<td>EPgV</td>
<td>OAS</td>
<td>GGTGCTGACACCGACAGTRGC</td>
</tr>
<tr>
<td>EPgV_NS5B_9222</td>
<td>EPgV</td>
<td>OS</td>
<td>AATTGAGTGAGCAGAGAAG</td>
</tr>
<tr>
<td>EPgV_NS5B_9231</td>
<td>EPgV</td>
<td>IS</td>
<td>CGCGGAGAAGATGATTCCTGGG</td>
</tr>
<tr>
<td>EPgV_NS5B_9666</td>
<td>EPgV</td>
<td>IAS</td>
<td>TACATCCCTGCTGATAAGGAAGGA</td>
</tr>
<tr>
<td>EPgV_NS5B_9780</td>
<td>EPgV</td>
<td>OAS</td>
<td>GCRGTATCCAGGAGWATGG</td>
</tr>
<tr>
<td>PanHCutr_OS</td>
<td>HCV</td>
<td>OS</td>
<td>CACWYCRCCATGAWRCACTCCCCCC</td>
</tr>
<tr>
<td>PanHCutr_IS</td>
<td>HCV</td>
<td>IS</td>
<td>TARCCATVRCSWAGTAYGRGTA</td>
</tr>
<tr>
<td>PanHCutr_IAS</td>
<td>HCV</td>
<td>IAS</td>
<td>GCMCTGCAGCAAGAYGTATCAG</td>
</tr>
<tr>
<td>PanHCutr_OAS</td>
<td>HCV</td>
<td>OAS</td>
<td>GGTCTACGACCKSMG</td>
</tr>
<tr>
<td>TDAV_NS3_OS</td>
<td>TDAV</td>
<td>OS</td>
<td>CAGACCGAGACTGATTGGG</td>
</tr>
<tr>
<td>TDAV_NS3_IS</td>
<td>TDAV</td>
<td>IS</td>
<td>GCTCTGAAAGGAAGAACGGC</td>
</tr>
<tr>
<td>TDAV_NS3_IAS</td>
<td>TDAV</td>
<td>IAS</td>
<td>TGTTAAGGCGATGGC</td>
</tr>
<tr>
<td>TDAV_NS3_OAS</td>
<td>TDAV</td>
<td>OAS</td>
<td>TCCGTACCCTAAGACTCCAC</td>
</tr>
<tr>
<td>EPgV_NPHV_DegOS</td>
<td>EPgV/NPHV</td>
<td>OS</td>
<td>TGCGRAGCTTTGTGGWRWRAG</td>
</tr>
<tr>
<td>EPgV_NPHV_DegIS</td>
<td>EPgV/NPHV</td>
<td>IS</td>
<td>TGCTACTGCTGACCCACCAGGG</td>
</tr>
<tr>
<td>EPgV_NPHV_DegIAS</td>
<td>EPgV/NPHV</td>
<td>IAS</td>
<td>GGCAATTYYGACACMGTCACYGAC</td>
</tr>
<tr>
<td>EPgV_NPHV_DegOAS</td>
<td>EPgV/NPHV</td>
<td>OAS</td>
<td>AAATTCCACMGTMACAYGCTACTG</td>
</tr>
<tr>
<td>GBV_NS3_Os</td>
<td>HPgV</td>
<td>OS</td>
<td>GSGCNATGGGNCNCTAYATGGA</td>
</tr>
<tr>
<td>GBV_NS3_IS</td>
<td>HPgV</td>
<td>IS</td>
<td>GTCGATNATHTGAYGATGTYCA</td>
</tr>
<tr>
<td>GBV_NS3_IAS</td>
<td>HPgV</td>
<td>IAS</td>
<td>TCRCAACTCMRCCCTGTKARTGCAAR</td>
</tr>
<tr>
<td>GBV_NS3_OAS</td>
<td>HPgV</td>
<td>OAS</td>
<td>GTNACYTCVACNACTCCCTCTCACC</td>
</tr>
<tr>
<td>HCVNS3a_4032s</td>
<td>Hepacivirus</td>
<td>OS</td>
<td>CCNACNGNAYGGNAAAGACG</td>
</tr>
<tr>
<td>HCVNS3b_4032s</td>
<td>Hepacivirus</td>
<td>IS</td>
<td>CCN CAN GGNYGGGNAAGTMAC</td>
</tr>
<tr>
<td>HCVNS3a_4678a</td>
<td>Hepacivirus</td>
<td>IAS</td>
<td>CCNTR TAN CCN GTC ATN AGG GCR TC</td>
</tr>
<tr>
<td>HCVNS3a_4678bb</td>
<td>Hepacivirus</td>
<td>OAS</td>
<td>CCN GTR TAN CCN GTA CAN AGG GCR TC</td>
</tr>
<tr>
<td>CHCV-XOF1</td>
<td>CHCV</td>
<td>OS</td>
<td>GGTGCCCTAYGACATYATITGGYGA</td>
</tr>
<tr>
<td>CHCV-XOF2</td>
<td>CHCV</td>
<td>IS</td>
<td>GCTAYGACATYATITGGYGA</td>
</tr>
<tr>
<td>CHCV-XOR1</td>
<td>CHCV</td>
<td>IAS</td>
<td>TACCTGTCAATGIGCC TCI GT</td>
</tr>
<tr>
<td>CHCV-XOR2</td>
<td>CHCV</td>
<td>OAS</td>
<td>ATGAGIGGCTCGTGICN</td>
</tr>
<tr>
<td>HBV49</td>
<td>HBV</td>
<td>OS</td>
<td>CCA ATT TGT CCT GGY TAT CG</td>
</tr>
</tbody>
</table>

21 a Orientation of primer sequences. OS : outer sense. OAS : outer antisense. IS : inner sense. IAS : inner antisense

22 Primer numbering is taken from the 5’ base position
List of Primers

<table>
<thead>
<tr>
<th>Primer</th>
<th>Type</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>HBV48</td>
<td>IS</td>
<td>CTG GAT GTG TCT GCG GCG TT</td>
</tr>
<tr>
<td>HBV50</td>
<td>IAS</td>
<td>TAA AGA GAG GTG CGK CCC GT</td>
</tr>
<tr>
<td>HBV51</td>
<td>OAS</td>
<td>GCA CAG ACG GGG AGA CCG CG</td>
</tr>
<tr>
<td>HBV52</td>
<td>OS</td>
<td>CWT TRT ATG CAT GTA TAC AAG C</td>
</tr>
<tr>
<td>HBV53</td>
<td>IS</td>
<td>TCG CCA AYT TAY AAG GCC TT</td>
</tr>
<tr>
<td>HBV54</td>
<td>IAS</td>
<td>GCG GTG TCR AGR AGA TCA CG</td>
</tr>
<tr>
<td>HBV55</td>
<td>OAS</td>
<td>GGC TTC MCG GTA CAR AGC TGA</td>
</tr>
<tr>
<td>HBV56</td>
<td>OS</td>
<td>TTT CCT KCT GAY TCC TTC CC</td>
</tr>
<tr>
<td>HBV57</td>
<td>IS</td>
<td>CGT GAT CTY CTY GAC ACC GC</td>
</tr>
<tr>
<td>HBV58</td>
<td>IAS</td>
<td>CAA GAA TAT GGT GAC CCA CA</td>
</tr>
<tr>
<td>HBV59</td>
<td>OAS</td>
<td>CCC ATG CTG TAG CTG TTG TTC C</td>
</tr>
<tr>
<td>HBV60</td>
<td>OS</td>
<td>ACA CAA TGT GGW TAT CCT GCM</td>
</tr>
<tr>
<td>HBV61</td>
<td>IS</td>
<td>GCC WTT GTA TGC ATG TAT ACA AGC</td>
</tr>
<tr>
<td>HBV62</td>
<td>IAS</td>
<td>GCC ACC CAA GGC ACA ACT TGG TG</td>
</tr>
<tr>
<td>HBV63</td>
<td>OAS</td>
<td>CTT GAA CAG TAG GAC ATG AAC AK</td>
</tr>
<tr>
<td>CHCVF1AK</td>
<td>CHCV</td>
<td>NS3 4203 TCC ACC TAT GGT AAG TTC TTA GC</td>
</tr>
<tr>
<td>CHCVF2AK</td>
<td>CHCV</td>
<td>NS3 4207 CCT ATG GTA AGT TCT TAG CTG AC</td>
</tr>
<tr>
<td>CHCVR1AK</td>
<td>CHCV</td>
<td>NS3 4635 ACC CTG TCA TAA GGG CGT C</td>
</tr>
<tr>
<td>CHCVR2AK</td>
<td>CHCV</td>
<td>NS3 4632 CCT GTC ATA AGG GCG TCC GT</td>
</tr>
<tr>
<td>CHCVF2SL</td>
<td>CHCV</td>
<td>NS3 4266 TGT GAT GAG TGC CATAGC ACD GA</td>
</tr>
<tr>
<td>CHCVR2SL</td>
<td>CHCV</td>
<td>NS3 4554 TAG TAG GTB ACA GCR TTA GGY CC</td>
</tr>
<tr>
<td>EQNS3OS</td>
<td>NPHV</td>
<td>NS3  ATW GTG GAT GAR TCR GAY AGY AC</td>
</tr>
<tr>
<td>EQNS3OAS</td>
<td>NPHV</td>
<td>NS3  TGY AAR GGT GTD AAG CTG TTG GT</td>
</tr>
<tr>
<td>EQNS3IAS</td>
<td>NPHV</td>
<td>NS3  TGG CAG AAG ATA AGR TGY CTY CC</td>
</tr>
<tr>
<td>EQNS3IS</td>
<td>NPHV</td>
<td>NS3  TAG TAG GTB ACA GCR TTA GGY CC</td>
</tr>
<tr>
<td>EQUROS</td>
<td>NPHV</td>
<td>5'UTR ACA YYY CCA TGT GTC ACT CCC CCT</td>
</tr>
<tr>
<td>EQUITRIAS</td>
<td>NPHV</td>
<td>5'UTR ACA CGG AAA YGG GTT AAC CAY ACY A</td>
</tr>
<tr>
<td>EQUITRIS</td>
<td>NPHV</td>
<td>5'UTR GCC CTC GCA AGC ATC TCA TCA G</td>
</tr>
<tr>
<td>EQUITROAS</td>
<td>NPHV</td>
<td>5'UTR CYC ATG TCC TAT GGT CTA CGA GA</td>
</tr>
<tr>
<td>CSUTROS</td>
<td>CHCV</td>
<td>5'UTR 90 TGT GTC ACT CCC CCT ATG GAG GGT TCC</td>
</tr>
<tr>
<td>CSUTRIS</td>
<td>CHCV</td>
<td>5'UTR 137 GTT TAA CCA TAC CCA AAG TAC GGG T</td>
</tr>
<tr>
<td>CSUTRIAS</td>
<td>CHCV</td>
<td>5'UTR 334 GCC CTC GCA AGC ATC TCA TCT GT</td>
</tr>
<tr>
<td>CSUTROAS</td>
<td>CHCV</td>
<td>5'UTR 369 TTA TTA TCT ATG TCC TAT GGT CTA CG</td>
</tr>
<tr>
<td>CSUTRIS</td>
<td>CHV</td>
<td>5'UTR 90 TGT GTC ACT CCC CCT ATG GAG GGT TCC</td>
</tr>
<tr>
<td>CSUTRIAS</td>
<td>CHV</td>
<td>5'UTR 137 GTT TAA CCA TAC CCA AAG TAC GGG T</td>
</tr>
<tr>
<td>CSUTROAS</td>
<td>CHV</td>
<td>5'UTR 334 GCC CTC GCA AGC ATC TCA TCT GT</td>
</tr>
<tr>
<td>CS5'UTROS</td>
<td>NPHV</td>
<td>5'UTR 81 ACA YYA CCA TGT GTC ACT CCC CCT</td>
</tr>
<tr>
<td>CS5'UTRIS</td>
<td>NPHV</td>
<td>5'UTR 126 ACA CGG AAA YGG GTT AAC CAY ACY A</td>
</tr>
<tr>
<td>CS5'UTROAS</td>
<td>NPHV</td>
<td>5'UTR 335 GCC CTC GCA AGC ATC TCA TCA G</td>
</tr>
<tr>
<td>CS5'UTROAS</td>
<td>NPHV</td>
<td>5'UTR 366 CYC ATG TCC TAT GGT CTA CGA GA</td>
</tr>
<tr>
<td>EQNS3OS</td>
<td>NPHV</td>
<td>NS3 4263 ATW GTG GAR TGC CAY AGY AC</td>
</tr>
<tr>
<td>EQNS3IS</td>
<td>NPHV</td>
<td>NS3 4335 TCY AAR GGT GTD AAG CTT GTT GT</td>
</tr>
</tbody>
</table>
EQNS3IAS  NPHV  NS3  4478  TGG CAG AAG ATA AGR TGY CTY CC
EQNS3OAS  NPHV  NS3  4554  TAG TAG GTB ACA GCR TTA GCY CC
EQNS5BOS  NPHV  NS5B  TGY ART ACA CTC CYC GYC ARC GTG T
EQNS5BIS  NPHV  NS5B  AAR TGY TTT GAC TCY ACB GTC ACT C
EQNS5BIS2 NPHV  NS5B  CAY GAT ATA GAH ACT GGT AGR GA
EQNS5BIAS NPHV  NS5B  ACT RTG ACT RAT YGT YTC CCA ACT CG
EQNS5BIAS2 NPHV  NS5B  TCR TCT TCC ACG CCY TTR CTG G
EQNS5BOS_8413 NPHV  NS5B  GAY CCR TAC ATC TCR AAC TCA TAA RT
EQNS5BIS_8437 NPHV  NS5B  1447 TCT TYA CDA ACG GGC AYT GGG GCG T
EQNS5BIS2_8466 NPHV  NS5B  1477 TTG CTC TWT GTC TYT ATG GCA CAA C
EQNS5BIAS_9031 NPHV  NS5B  2325 CAC TGT ARA GHA GGG AGA CGA TAC T
EQNS5BOAS_9031 NPHV  NS5B  2430 CCC AMA CAT ACT GYA CRA TAC TCA GA
EQNS5BOS_8413 NPHV  NS5B  2101 CAA CTN TGT GGC CAR GGA CCG TG
EQNS5BIS_8437 NPHV  NS5B  2286 GAH AGG TTT AGA GGG ACT AAT GTC C
EQNS5BIS2_8466 NPHV  NS5B  3134 ACA CCA CTG HAC CCA YTT RCT ACC
EQNS5BIAS_9031 NPHV  NS5B  3173 ATADAGGAAAGT VCC AAR RTG AGA AGC
EQNS5BOAS_9031 NPHV  NS5B  3009 DTC CTT TTC CAT GAT CAA CAA GCG YG
EQNS5BIS_8437 NPHV  NS5B  2891 GCT TGG CGC TAT CTC GGC CCT AGG GT
EQNS5BIS2_8466 NPHV  NS5B  3063 CGC GCB GAG GTC TTG GTK MAG GT
EQNS5BIAS_9031 NPHV  NS5B  2725 GAY CCR TAC ATC TCR AAC TCA TAA RT
EQNS5BOAS_9031 NPHV  NS5B  4044 GCA TTY TGG TGG AYT TGC CRC TRC
EQNS5BIS_8437 NPHV  NS5B  4165 CCM GTC CTGATG TTA GGG CAY TCM CC
EQNS5BIS2_8466 NPHV  NS5B  4281 GTR CTR TGGCAY TCA TCA CAW AT
EQNS5BIAS_9031 NPHV  NS5B  4646 TGACAG GGTATT CTG GCA ATT TCG ACA
EQNS5BOAS_9031 NPHV  NS5B  4761 CTGGACATTGAATTT TCYCTTGACCCAC
EQNS5BIS_8437 NPHV  NS5B  5762 GGCAAWCCRTGAGA CCW GAA ACA
EQNS5BIS2_8466 NPHV  NS5B  5744 GAATGTTGAGGA AAG CCA TAG CTG GHC
EQNS5BOAS_9031 NPHV  NS5B  6275 TTA TGACTG TTT CAC CTG GGC AAA GGG
EQNS5BIS_8437 NPHV  NS5B  7265 GGGCTT GAT GGG ACA CCC GGA WAC GA
EQNS5BIS2_8466 NPHV  NS5B  7286 ARGGCT CAC TGG TGY CAG GVG TAG CA
EQNS5BOAS_9031 NPHV  NS5B  7106 TCAATTCTG GTT CNG AGG GRG AGC ACT
EQNS5BIS_8437 NPHV  NS5B  7203 TGG ATGCGV AAA GAC TAY TGC CCA GC
EQNS5BIS2_8466 NPHV  NS5B  8197 ATG ACTGTC TTA GGG CAY TCM CC
EQNS5BOAS_9031 NPHV  NS5B  8290 TCTCCT TCC AAC CAG GTC CCC ACG GT
EQNS5BIS_8437 NPHV  NS5B  5481 CAA GGD TAY GAR TGG GCG TCG CG
EQNS5BIS2_8466 NPHV  NS5B  5583 ARTTTT GGAATC ACA GCA TGT GGA ACA
EQNS5BOAS_9031 NPHV  NS5B  5551 CTGGTCAAACAACCCWGTGATCRG C
EQNS5BIS_8437 NPHV  NS5B  6469 CGGCAC TTC CGC GAG CCT GTD AT
EQNS5BIS2_8466 NPHV  NS5B  6502 GARTTRATTGGG AAR GYA TTG AGC CAG
EQNS5BOAS_9031 NPHV  NS5B  8876 TTACGAAGTTAAG AGT GGC TGC AGC
EQNS5BIS_8437 NPHV  NS5B  8827 TTT ATA AAC TTC ATG GCC CAG AGG GC
EQNS5BIS2_8466 NPHV  NS5B  8827 TTT ATA AAC TTC ATG GCC CAG AGG GC
EQNS5BOAS_9031 NPHV  NS5B  5'UTR1  CCT CCG TGC TAG GCA CCG TGC GT
EQNS5BIS_8437 NPHV  NS5B  5'UTR7  TGC TAG GCA CCG TGG GTT GTC AGC G
EQNS5BOS_8413 NPHV  NS5B  CCA GYA ARG GCG TYG AGG AAG AYG A
EQNS5BIS_8437 NPHV  NS5B  ACA GGC GCT GCG AGA TTT CAC CA
EQNS5BIS2_8466 NPHV  NS5B  ATG ACA AAG TAC TCC GCC CTT CCT GG
EQNS5BIAS_9031 NPHV  NS5B  GCA GAA GCG AAG CCR AAR AGR TAG TC
EQNS5BOAS_9031 NPHV  NS5B  GAG TCA TCT CCA ATG GAG AAC CAA
Appendix B – Supplementary data in the determination of EPgV 5’UTR secondary structure

Figure B1.1 PFOLD RNA secondary structure analysis of the 5’UTR of EPgV from nucleotide position 1-400 (Fragment 1). Each nucleotide pairing corresponds to coordinates [x, y] on the DotPlot to generate map of predicted stem-loop structures.
Figure B1.2 PFOLD RNA secondary structure analysis of the 5'UTR of EPgV from nucleotide position 290-627 (Fragment 1). Corresponds to the overlapping fragment for Figure A.1.1 to complete the 5'UTR structural prediction. Each nucleotide pairing corresponds to coordinates \([x, y]\) on the DotPlot to generate map of predicted stem-loop structures.
Figure B1.3 ALIFOLD RNA secondary structure prediction of the 5’UTR of EPgV from nucleotide position 1-630 (Fragment 1). ALIFOLD based prediction of stem-loops and secondary structural elements in the 5’UTR of EPgV. Each nucleotide pairing corresponds to coordinates \([x, y]\) on the DotPlot to generate map of predicted stem-loop structures.
Supplementary data for EPgV 5′ UTR secondary structure

Figure B1.4 ALIFOLD RNA interactive model of the predicted 5′ UTR of EPgV from nucleotide position 1-627. Colour coding of the interactive model predicts the likelihood of nucleotide base pairings occurring, scale 0 to 1.
Appendix C: Permission to reproduce figures

This is a License Agreement between Sinead Lyons ("You") and Elsevier ("Elsevier") provided by Copyright Clearance Center ("CCC"). The license consists of your order details, the terms and conditions provided by Elsevier, and the payment terms and conditions.

**Supplier**
Elsevier Limited The Boulevard, Langford Lane Kidlington, Oxford, OX5 1GB, UK

**Registered Company Number**
1982084

**Customer name**
Sinead Lyons

**Customer address**
University of Edinburgh, Edinburgh, EH25 9RG

**License number**
3353281182017

**License date**
Mar 20, 2014

**Licensed content publisher**
Elsevier

**Licensed content publication**
Vaccine

**Licensed content title**
Hepatitis B virus genotypes

**Licensed content author**
Anna Kramvis, Michael Kew, Guido François

**Licensed content date**
31 March 2005

**Licensed content volume number**
23

**Licensed content issue number**
19

**Number of pages**
15

**Start Page**
2409

**End Page**
2423

**Type of Use**
reuse in a thesis/dissertation

**Intended publisher of new work**
other

**Portion**
figures/tables/illustrations

**Number of figures**
2

**Format**
both print and electronic

**Are you the author of this Elsevier article?**
No

**Will you be translating?**
No

**Title of your thesis/dissertation**
Evolutionary history, cross-species transmission and host adaptation of human viruses and their primate homologues

**Expected completion date**
Apr 2014

**Estimated size (number of pages)**
350

**Elsevier VAT number**
GB 494 6272 12

**Permissions price**
0.00 USD
This is a License Agreement between Sinead Lyons ("You") and Springer ("Springer") provided by Copyright Clearance Center ("CCC"). The license consists of your order details, the terms and conditions provided by Springer, and the payment terms and conditions.

License Number: 3353281489409
License date: Mar 20, 2014
Licensed content publisher: Springer
Licensed content publication: Springer eBook
Licensed content title: The Origin of Hepatitis C Virus
Licensed content author: Peter Simmonds
Licensed content date: Jan 1, 2013
Type of Use: Thesis/Dissertation
Portion: Figures
Author of this Springer article: No
Original figure numbers: Fig. 1.
Title of your thesis / dissertation: Evolutionary history, cross-species transmission and host adaptation of human viruses and their primate homologues
Expected completion date: Apr 2014
Estimated size (pages): 350
Total: 0.00 USD
Appendix D: Transcript Sequences

A) NPHV

NS3

GGCGCTTATGACATAATCATCTGTGATGAGTGCCATAGCACAGATTCCAC
GTCCGTTTAGGAATGGCTCTGTCTGGTCTGGACGGGGGGAGTCTAGGG
TTAAGCTTGTTTGGTGTGCTACTGCCAACCAGGGTCCAATAACTGTTTC
CTCATCCAAACATCGACGAGGAGGCCCTTACTCAGAGTGAGACATCCC
CTTCTAGGGGAAGATGTTGAAGTGCCTTGCTCTAFCGTTGAGTGGAGTG
TTATCTT

TTTGCCATTCGACTGAAGAAGAAGTGTGAGGAGCCTGCTCCCTTTTG
AGAAAGGCTGAGCTAATGCTGACACTACCCGGGGCTTTAGGATGTTTC
CGTCATTCCGAATGAGGGGAATGTTGTCGTTGTTGCCACAGATGCCCTCA
TGACAGGGT

B) EPgV

NS3

GAGATACCCTGACTGCTGAGGGAGACATTCCCTTTATGCTTTAAAAT
CCCCTGGCTGAAAATTTACGAGAAGGGAGACATTTGTTTTTGTCCATTCA
AAAGGCTGAGCATAAGACTGCACACTCCTTTACACAGACATGGCTGCTG
GGCAATTACACTATATTCTGGTCGAGATACGCTCCAGATCCCTCCTTT
Appendix E: Publications

Published

(A)

(B)

Accepted (April 2014)

(C)
Species Association of Hepatitis B Virus (HBV) in Non-Human Apes; Evidence for Recombination between Gorilla and Chimpanzee Variants

Sinéad Lyons, Colin Sharp, Matthew LeBreton, Cyrille F. Djoko, John A. Kiyang, Felix Lankester, Tafon G. Bibila, Ubald Tamoufe, Joseph Fair, Nathan D. Wolfe, Peter Simmonds

1 Centre for Immunology, Infection and Evolution, University of Edinburgh, Edinburgh, United Kingdom, 2 Roslin Institute, University of Edinburgh, Midlothian, United Kingdom, 3 Limbe Wildlife Centre, Limbe, Cameroon, 4 Global Viral Forecasting, San Francisco, California, United States of America, 5 Global Viral Forecasting, Stanford University, Program in Human Biology, Stanford, California, United States of America, 6 Ape Action Africa, Yaounde, Cameroon

Abstract

Hepatitis B virus (HBV) infections are widely distributed in humans, infecting approximately one third of the world’s population. HBV variants have also been detected and genetically characterised from Old World apes; Gorilla gorilla (gorilla), Pan troglodytes (chimpanzee), Pongo pygmaeus (orang-utan), Nomascus nasutus and Hylabates pileatus (gibbons) and from the New World monkey, Lagotrichia lagotrichia (woolly monkey). To investigate species-specificity and potential for cross species transmission of HBV between sympatric species of apes (such as gorillas and chimpanzees in Central Africa) or between humans and chimpanzees or gorillas, variants of HBV infecting captive wild-born non-human primates were genetically characterised. 9 of 62 chimpanzees (11.3%) and two from 11 gorillas (18%) were HBV-infected (15% combined frequency), while other Old world monkey species were negative. Complete genome sequences were obtained from six of the infected chimpanzee and both gorillas; those from P. t. ellioti grouped with previously characterised variants from this subspecies. However, variants recovered from P. t. troglodytes HBV variants also grouped within this clade, indicative of transmission between sub-species, forming a paraphyletic clade. The two gorilla viruses were phylogenetically distinct from chimpanzee and human variants although one showed evidence for a recombination event with a P.t.e.-derived HBV variant in the partial X and core gene region. Both of these observations provide evidence for circulation of HBV between different species and sub-species of non-human primates, a conclusion that differs from the hypothesis if of strict host specificity of HBV genotypes.


Introduction

Hepatitis B virus (HBV) is a member of the Hepadnaviridae family of viruses, containing a partially double-stranded DNA genome of approximately 3182–3221 nucleotides [1]. Human hepatitis B virus is globally distributed, infecting approximately one third of the world’s human population. A substantial proportion of liver disease is attributable to HBV, killing over one million people each year [2]. In South and East Asia, Sub-Saharan Africa and South and Central America populations show a particularly high frequency of HBV infection which can be maintained by vertical transmission between mother to child transmission or horizontal transmission during childhood [3].

HBV variants infecting humans show genetic and antigenic heterogeneity and are currently classified into 7 or 8 genotypes (A–H) with a nucleotide sequence divergence ranging from 9% to 13%. Two putative genotypes I and J have also been reported. Genotype I was tentatively suggested for strains recovered in Laos [4]. A ninth genotype J was recovered from an 88-year-old Japanese patient with hepatocellular carcinoma, with mean sequence divergence between HBV/J and gibbon and orangutan genotypes of 10.9% and 10.7% respectively [5]. Both active and resolved HBV infections are also found at high frequencies in chimpanzees [6,7] and South Asian apes [7]; whose 10–12 host taxa-associated variants are distinct from the human variants of the virus. In addition to the current HBV genotypes, recombination between human genotypes, for example between genotypes A and D [1,8,9,10] and B and C [8,9,11,12] can generate novel variants, contributing to the genetic diversity of the virus.

Within the past 10 years, both active and resolved HBV infections have also been found in chimpanzees [2], gorillas [13], gibbons [13,14] and orang-utans [7] at infection frequencies comparable to human rates in endemic regions [15,16] in addition to a single isolate from a woolly monkey [17]. Furthermore, tentative evidence for the occurrence of recombination has been obtained between the human genotype C and the chimpanzee...
variant AF498266 [18] and gibbon variants [15]. Their occurrence demonstrates that despite their genetic divergence, human and non-human associated variants of HBV can share hosts in nature. A recently published study characterising HBV variants infecting ape populations in Cameroon [21] demonstrated the existence of a gorilla-specific HBV strain and evidence of recombination between HBV strains circulating in chimpanzees. This and previous studies of HBV nucleotide sequence similarity [14] indicate non-human primates (NHP) have distinct species-specific variants of HBV distinguishable both from each other and from human HBV despite occupying overlapping geographical areas.

Cameroon is within a region of endemic human HBV infection with a hepatitis B surface antigen (HBsAg) prevalence in humans of 8% or greater [20]. Additionally, four different great ape taxa also occur in Cameroon, providing the conditions for potential inter-species transmission. Although no human-derived genotypes of HBV were detected in non-human primates in the current study, evidence for transmission of HBV between chimpanzee subspecies and between chimpanzees and gorillas was obtained.

Results

A total of 164 non-human primate plasma samples from 11 gorillas, 62 chimpanzees and 91 Old World Monkeys (OWM) were screened for the presence of HBV DNA. PCR screening showed 9/75 (12%) apes were positive corresponding to 2/11 gorillas (18%) and 7/62 chimpanzees (11.5%) and 2/91 (2.2%) OWM. Complete HBV genomes were obtained from the isolates of 2 Gorilla gorilla and 6 chimpanzeas (4 P. t. ellioti, 2 P. t. troglodytes), while both the Old World Monkey isolates (1 Grey cheeked mangabey and 1 Mandrill) and 1 chimpanzee were positive only with the screening primers originally used.

Phylogenetic analysis of the HBV strains using 415 bp S gene fragments confirmed the grouping of the novel chimpanzee HBV strains with previously published chimpanzee HBV sequences and the grouping of the two novel gorilla HBV sequences with previously published HBV strains AJ131657 and FJ98095-97 [19,21]. The novel HBV sequence ECO50003LIP3 and FJ98095 were retrospectively identified as originating from the same gorilla in Limbe Wildlife Centre (Data not shown). Mitochondrial sequencing confirmed that the ECO50083LIP5, ECO50210LIP4, ECO51394CW1.4, ECO51377CW2 and ECO51109CW4, ECO51212CW6 HBV variants originated from chimpanzee subspecies P. t. ellioti and P. t. troglodytes respectively, while ECO50003LIP3 and ECO50065LIP3 were identified as gorilla-derived (Table 1). Complete genome sequencing of the eight study isolates produced sequences of 3182-bp in length, comparable to reference chimpanzee and gorilla strains (Fig 1). Phylogenetic analysis based on the complete genome (Fig. 1a), demonstrated monophyletic groupings for each human genotype (A-H), a clade containing gibbon and orangutans variants and a third containing chimpanzee and gorilla HBV sequences, each supported by high bootstrap values. Sequences of all novel chimpanzee HBV variants, irrespective of their sub species specific host, clustered with HBV sequences previously obtained from P. t. ellioti [21]. As the study samples were obtained from captive settings where P. t. troglodytes and P. t. ellioti are frequently co-housed, it remains unclear whether these findings reflect the HBV genotype distribution among wild chimpanzees in Cameroon or whether infection of troglodytes by ellioti-derived strains occurred in captivity.

Phylogenetic trees of 200 bp fragments incrementing by 50 bp across the entire genome were constructed to identify changes in phylogeny potentially indicative of recombination events. This procedure was automated by the program TreeOrder Scan [22] in SSE v1.0 [Manuscript in preparation]. The tree position of each sequence across the genome is recorded (y-axis) and colour-coded by HBV genotype and species (Fig. 2a) and by sub-species (Fig. 2b). Changes in the tree order of individual sequences or genotypes with 70% or greater bootstrap support are indicative of alterations in the phylogenetic relationships of clades and identify potential recombination breakpoints.

Consistent with previous findings [22], phylogenetic relationships between human genotypes changed between genome regions, leading to alterations in the branching order. For example, genotypes D and E were largely phylogenetically distinct with 70% or greater bootstrap support in the gorilla and chimpanzee sequences, recombination events typically occur around positions 730, 900, 1600, 1950, 2500, 2650 and 3182.

Table 1. Specimen isolation data.

<table>
<thead>
<tr>
<th>Specimen Number</th>
<th>Mito</th>
<th>HBV Variant</th>
<th>Location</th>
<th>Date arrival</th>
<th>Approx age on arrival</th>
<th>Date sample collection</th>
<th>Previous holder</th>
<th>Likely Wild Origin</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gorilla ECO50003LIP3</td>
<td>G.g</td>
<td>Gorilla gorilla</td>
<td>Limbe Wildlife Centre</td>
<td>13-Sep-03</td>
<td>9 years</td>
<td>19-Aug-04</td>
<td>Nikoma, Mvangan, South Region</td>
<td>South Region</td>
</tr>
<tr>
<td>Gorilla ECO50065LIP3</td>
<td>G.g</td>
<td>Gorilla gorilla</td>
<td>Limbe Wildlife Centre</td>
<td>17-Dec-96</td>
<td>9 months</td>
<td>22-Jun-05</td>
<td>Bertoua, East Region</td>
<td>East Region</td>
</tr>
<tr>
<td>Chimpanzee ECO50083LIP5</td>
<td>P.t.e.</td>
<td>Pan troglodytes ellioti</td>
<td>Limbe Wildlife Centre</td>
<td>04-Oct-05</td>
<td>1.5 years</td>
<td>12-Oct-05</td>
<td>Garoua Zao (North Region)</td>
<td>Adamaoua or East Region</td>
</tr>
<tr>
<td>Chimpanzee ECO50210LIP4</td>
<td>P.t.e.</td>
<td>Pan troglodytes ellioti</td>
<td>Limbe Wildlife Centre</td>
<td>31-Dec-04</td>
<td>8 months</td>
<td>05-Dec-06</td>
<td>Bachou, Manyu, South West Region</td>
<td>Banyang Mbo Wildlife Sanctuary</td>
</tr>
<tr>
<td>Chimpanzee ECO51109CW4</td>
<td>P.t.t.</td>
<td>Pan troglodytes ellioti</td>
<td>Mfou NP sanctuary</td>
<td>25-May-00</td>
<td>1.25 years</td>
<td>08-Aug-06</td>
<td>Mfou, Centre Region</td>
<td>Akom II, South Region</td>
</tr>
<tr>
<td>Chimpanzee ECO51212CW6</td>
<td>P.t.t.</td>
<td>Pan troglodytes ellioti</td>
<td>Mfou NP sanctuary</td>
<td>08-Aug-06</td>
<td>1 years</td>
<td>10-Mar-08</td>
<td>Unknown</td>
<td>East Region</td>
</tr>
<tr>
<td>Chimpanzee ECO51394CW1.4</td>
<td>P.t.e.</td>
<td>Pan troglodytes ellioti</td>
<td>Mfou NP sanctuary</td>
<td>13-Sep-05</td>
<td>1.5</td>
<td>26-Aug-09</td>
<td>Unknown</td>
<td>Centre Region</td>
</tr>
<tr>
<td>Chimpanzee ECO51377CW2</td>
<td>P.t.e.</td>
<td>Pan troglodytes ellioti</td>
<td>Mfou NP sanctuary</td>
<td>01-Oct-97</td>
<td>4.6 yrs</td>
<td>26-Jul-09</td>
<td>Unknown</td>
<td>Unknown</td>
</tr>
</tbody>
</table>
Figure 1. Phylogenetic analysis based on the HBV genome and identified recombinant region 1560–2120 bp. Phylogenoms displaying phylogenetic trees based on (a) complete HBV genome; (b) HBV recombinant region 1560–2120 bp and equivalent fragments immediately preceding; (b) 999–1559 bp and succeeding; (d) 2121–2681 bp this region; with HBV reference sequences from human genotypes (A–H). Relative species and sub-species HBV variants are identified as follows Pan troglodytes schweinfurthii; Pan troglodytes ellioti; Pan troglodytes verus; Pan troglodytes schweinfurthii; Gorilla gorilla; and Hylobates spp. The host specific cluster is identified by . The trees were rooted with the woolly monkey HBV sequence, NC_001896. Sequences from this study are in bold and underlined and while recombinant HBV variants have bold branches.

2750, frequently coinciding with gene boundaries as previously described [22].

In order to detect recombination events between HBV variants from different ape taxa, a Tree Order Scan was performed with all HBV reference and study sequences from gorillas and chimpanzees with a Hylobates pileatus sequences as an out-group (Fig. 2b). Outside of the core gene region, gorilla-derived variants were phylogenetically distinct from other ape-associated and human genotypes. However, between positions 1560 and 2120 (in the X and pre-core gene), the gorilla isolate ECO30003LIP3 sequence grouped within the chimpanzee clade, indicative of a recombination event. ECO30003LIP3 showed 99.45% (99.0–99.9%) similarity with other gorilla variants between positions 1 and 1559 and 99.65% (99.5–99.8%) similarity between positions 2121–3272, and an average 95.3% and 97.4% similarity with P.ellioti variants across the same respective regions.

This analysis also identified recombination in the sequence A498266 (Fig 2b), a P. t. schweinfurthii HBV isolate previously identified as a recombinant between human genotype C and chHBV [18]. For this sequence, a 500 nt region between positions 550–1050 nt grouped with species C while the remainder of the genome grouped with P.troglodytes sequences, consistent with previous findings [18]. The recently described HBV sequence from P.troglodytes [19] may therefore represent the “original” P. t. schweinfurthii sequence, from which the recombinant arose [19]. The inclusion of these sequenced isolates described in our investigation supports the formation of a P. t. schweinfurthii species-specific clade (Fig. 1) which includes the genotype C/chHBV recombinant. Corroborating evidence of recombination was identified in Cameroon chimpanzee sequences, FJ90898.1 and FJ90899.1, between positions 820 and 1300 nt, traversing a partial specific clade (Fig. 1) which includes the genotype C/chHBV [19]. The inclusion of these sequenced isolates described in our previous findings [18]. The recently described HBV sequence grouped with species C while the remainder of the chHBV [18]. For this sequence, a 500 nt region between positions 550–1050 nt, traversing a partial specific clade (Fig. 1) which includes the genotype C/chHBV [19]. The inclusion of these sequenced isolates described in our previous findings [18]. The recently described HBV sequence grouped with species C while the remainder of the core gene region, gorilla-derived variants were phylogenetically distinct from other ape-associated and human genotypes. However, between positions 1560 and 2120 (in the X and pre-core gene), the gorilla isolate ECO30003LIP3 sequence grouped within the chimpanzee clade, indicative of a recombination event. ECO30003LIP3 showed 99.45% (99.0–99.9%) similarity with other gorilla variants between positions 1 and 1559 and 99.65% (99.5–99.8%) similarity between positions 2121–3272, and an average 95.3% and 97.4% similarity with P.ellioti variants across the same respective regions.

To confirm the position and phylogenetic grouping of the putative recombinant sequences identified by the TreeOrder scan, a Grouping Scan was performed [22]. This examines how deeply the embedded test sequence is within clades formed by non-recombinant control sequences assigned into species-associated groups (P.t.e, P.t.s, P.t.g and Gorilla gorilla) (Fig. 3). This method identified two changes in grouping of the query sequence, ECO30003LIP3 at position 1560 where it changed grouping from the gorilla HBV clade to the P. t. ellioti sub-species clade and a reversion to the gorilla clade at position 2120. Grouping scan analysis of recombinant sequences A498266 and FJ90898.1 provides substantial support for the formation of recombinant regions between positions 550–1050 nt and 820–1300 nt respectively. However, the Tree Order or grouping scan methods provided no evidence for recombination in the P.troglodytes derived sequence, AM117396 (based on its grouping in Fig. 1a) between chimpanzee sub-species.

Sequence AB046525 from P.troglodytes in Central Africa grouped separately from other P.tl variants in core gene region, consistent with past recombination with a divergent and currently uncharacterised genotype of HBV [23]. Both TreeOrder scan analysis and Grouping Scan analysis confirmed the rest of the genome groups consistently with P.troglodytes, while adopting an outlier position to all other chimpanzee and gorilla isolates in the core region.

Discussion

In this study a large number plasma samples from great apes and monkeys from Cameroon were screened for HBV-DNA. The prevalence amongst chimpanzees in our study was found to be 9.7% (6/62), 18% (2/11) in gorillas. This confirms previous findings on the existence of HBV in great apes in the wild [16,22] and the rates are similar to those observed in human populations in areas of endemic infection, such as Central Africa and South East Asia. In a recent study, the prevalence of active HBV infection (DNA-positive in plasma) was 15% (8/53) in gorillas and 18% (40/203) in chimpanzees [15,24].

Eight new complete HBV genomes were obtained in the current study from two gorillas and six chimpanzees born in the wild. Gorilla sequence ECO300065LIP3 was almost identical to the previously described sequence, FJ798095 [21], and retrospective analysis revealed that these sequences originated from the same animal in Limbe Wildlife Centre. The six complete chimpanzee HBV sequences all grouped with previously identified P.ellioti variants [21] although two were recovered from P.troglodytes. Current analyses cannot determine whether these two cross species infections occurred in the wild or through contact with infected P.ellioti chimpanzees while in captivity although the latter is highly likely given the mixing of chimpanzee subspecies in sanctuaries. The existence of P.troglodytes and P.ellioti associated variants of HBV, as is the case for other chimpanzee subspecies (P.troglodytes and P. schweinfurthii) requires further investigation of variants infecting chimpanzees in the wild in Cameroon, in particular in regions where these sub-species may converge, for example around the confluence of the Mbam and Sanaga Rivers [23]. The observation of cross-species infections and recombination events for HBV infections also provides an additional reason for ensuring that captive chimpanzees are correctly identified to subspecies and segregated appropriately to avoid the creation of recombinant HBV variants with potentially different pathogenicities and transmission patterns.

The phylogenetic tree comparing these eight sequences to previously recorded HBV sequences in non-human primates confirms that variants recovered from chimpanzees and gorillas in Africa are distinct from those reported in Asian gibbons and from all human HBV sequences that cluster separately into genotypes A–H. This is consistent with the geographical association of HBV in NHP previously reported [24]. Phylogenetic analysis based on complete genome found significant bootstrap support for the formation of four HBV clusters than, excluding likely cross-species transmissions, corresponded with P. troglodytes sub-clades: P. t. troglodytes (81% bootstrap), P. t. verus (100% bootstrap), P. t. ellioti (100% bootstrap) and P. t. schweinfurthii (100% bootstrap). Recent phylogenetic analysis of the P. t. schweinfurthii HBV strain; confirmed by our GroupScan analysis (Fig. 3c) showed evidence of interspecies recombination between HBV infecting chimpanzees and the human HBV-C genotype strain [6,18]. Phylogenetic trees of the recombinant region and equivalent fragments either
Figure 2. Tree Order Scan of HBV sequences. Figure 2(a). TreeOrder Scan of HBV sequences, indicating positions of individual sequences (y axis) in phylogenetic trees generated from sequential 250-base sequence fragments, incorporating by 50 bases. Changes in sequence order as a result of changes in phylogeny at the 70% bootstrap level are shown. Sequences are colour coded by genotype and host species, as indicated by the labels in left and right margin: genotype A: purple; B: light blue; C: wine; D: emerald; E: royal blue; F: orange; G: navy; Gorilla; blue (Gor); Chimpanzee, green (Pan); and Woolly monkey (WM-out-group on line 1). red. For comparison the Tree Order Scan has been aligned with scale genome of HBV (top panel). Recombinant sequences are highlighted as by dashed lines; black gorilla/P.t.e ECO50033LIP3, green FJ798099 P.t.e/P.Ltt, pink FJ798098 P.t.e/P.Ltt, orange AB046525 P.t.t and purple AF498266 P.t.e. For comparison the Tree Order Scan has been aligned with scale genome of HBV (top panel). Recombinant sequences are highlighted as by dashed lines; black gorilla/P.t.e ECO50033LIP3, green FJ798099 P.t.e/P.Ltt, brown FJ798098 P.t.e/P.Ltt, orange AB046525 P.t.t and blue AF498266 P.t.t.

doi:10.1371/journal.pone.0033430.g002

For comparison the Tree Order Scan has been aligned with scale genome of HBV (top panel). Recombinant sequences are highlighted as by dashed lines; black gorilla/P.t.e ECO50033LIP3, green FJ798099 P.t.e/P.Ltt, brown FJ798098 P.t.e/P.Ltt, orange AB046525 P.t.t and blue AF498266 P.t.t.

doi:10.1371/journal.pone.0033430.g002

side, inclusive of all reference and study sequences, confirm the sub-species association of HBV in NHPs [Fig. 1b to 1d]. The phylogenograms also support the recombinant data of the Tree Order and Grouping scan analysis, with respect to the location and confidence level for the recombinant region and sequence (Fig. 1b). The correlation of HBV sequences with the different sub-species of chimpanzees indicates either that the HBV strains and their hosts have co-evolved or alternatively have diverged through allopatric separation.

The co-divergence hypothesis for the distribution of non-human HBV genotypes in Africa and South East Asia presupposes that the distinct variants of HBV found in different ape species and subspecies arose during the period of their evolution, over the past 5-10 million years. However, such a hypothesis implies an extraordinarily slow maintained substitution rate of HBV, the 5% divergence between gorilla and chimpanzee variants indicates that the HBV strains and their hosts have co-evolved or alternatively have diverged through allopatric separation.

The co-divergence hypothesis for the distribution of non-human HBV genotypes in Africa and South East Asia presupposes that the distinct variants of HBV found in different ape species and subspecies arose during the period of their evolution, over the past 5-10 million years. However, such a hypothesis implies an extraordinarily slow maintained substitution rate of HBV, the 5% divergence between gorilla and chimpanzee variants indicates that the HBV strains and their hosts have co-evolved or alternatively have diverged through allopatric separation.

For comparison the Tree Order Scan has been aligned with scale genome of HBV (top panel). Recombinant sequences are highlighted as by dashed lines; black gorilla/P.t.e ECO50033LIP3, green FJ798099 P.t.e/P.Ltt, brown FJ798098 P.t.e/P.Ltt, orange AB046525 P.t.t and blue AF498266 P.t.t.

doi:10.1371/journal.pone.0033430.g002

Recombination affecting a short region between either end of the polymerase gene (partial X gene and Pre-core/core) in one of the gorilla isolates is the recombination event between chimpanzee- and gorilla-derived HBV variants. The recombination event between the gorilla and P.t.ellioti variants likely occurred in the wild as gorillas and chimpanzees are never co-housed in captivity. The position of the breakpoint region is close to several documented previously in human genotypes [22] (Fig. 2a). A recent study of duck HBV [31] recorded a similar breakpoint event between position 1010–2304 bp, incorporating the region of the X gene, which is believed to promote cell growth and inactivate growth regulating molecules [32,33].

Complex epidemiological factors such as transmission routes affect the pool of circulating HBV variants; however their spread may be enhanced by the evolution of recombinant variants, allowing the virus to transmit more efficiently between species. Phylogenetic studies have previously indicated that recombination events in HBV are quite common [12,22] and recombinant strains have been shown to possess distinct biological features and produce different clinical outcomes compared to their parental strains [31]. Further work is required to investigate the distribution of HBV recombinants in Cameroon, their potential impact on host species and the evolution of HBV in NHPs in the wild.

The evidence of animal recombinants, cross-species transmission and recombination between human and ape HBV variants have
Samples were screened by nested PCR using first-round primers according to the manufacturer’s instructions with DNA eluted in a final volume of 50 μl. Sequences of genotypes A–H. PCR positive samples with screening primers included samples from 9 apes (n = 2 gorillas, n = 7 chimpanzees) and 2 Old World Monkeys. The entire HBV genome was successfully sequenced for 2 gorillas (ECOS00003;IP3 and ECOS00065) and 6 chimpanzees (ECOS00083, ECOS00210, ECOS51109, ECOS51212, ECOS51394, AND ECOS51377) in overlapping fragments using primers as published [2] in addition to new primer sets: Set 1: Outer Sense 49 (CTGGATGTGTCTCGCCGGTT) position 375 and Outer Anti-Sense 51 (GCACAGCAGGGAGACCGCG position 1542) followed by Inner Sense 48 (CCAAATTGCTGCTGGTATCG position 395) and Inner Anti-Sense 50 (TAAAGAGGGTGC-GCCCGG position 1332), Set 2: Outer Sense 52 (CWTR- TATGGATGTATCAAGGCG position 1082) and Outer Anti-Sense 53 (GGCTTCMCGTACARAGCTGTA position 2052) followed by Inner Sense 53 (TGCGCAAYTTAYAAGGCCCT position 1121), and Inner Anti Sense 54 (GGGTTGCARAGRACGAC position 2032), and finally Set 3: Outer Sense 56 (TTCGCTTGAAYTTCTTCG position 2025) and Outer Anti-Sense 59 (CCCGTGCTGCTGCCATGTTCC position 2080), followed by Inner Sense 57 (CGGATCCTYCTYGAGACCGG position 2052) and Outer Anti-Sense 58 (CAAGATTATGTTGACCCCA position 2080). First round PCR involved 35 cycles of 94°C for 18 s, 55°C for 21 s, and 72°C for 1 min; and 1 cycle of 72°C for 5 min, followed by second round PCR of 40 cycles at matching conditions. Touch-down PCR between 65°C and 50°C with 0.5°C decline per cycle was also applied.

Mitochondrial sequencing to confirm host species and sub-species for the 9 complete genomes was carried out using primate specific primers: Forward: PrCOI CTATTYGGYGGATAGGCG- NGG Reverse: PrCOI TARAAGAARGTRTRTTRATGG- TRC, followed by Forward: PrCOI CAGGCTAAGYCTYC- TYATGG and Reverse: PrCOI GAYDGTACAGACRA- YARGG [Where R: A/G, Y: T/C, D: G/A/T and N: G/A/ T/C]. First and second round PCR conditions involved 30 cycles of 94°C for 22 s, 50°C for 24 s, 72°C for 1.5 min; and 1 cycle of 72°C for 5 min. Sequencing of PCR products and sequence analysis

Positive second round PCR amplicons were sequenced in both directions using the inner sense and inner antisense primers used in the second round of amplification. Sequencing was executed using Big Dye Terminator v3.1 (Applied Biosystems) according to the manufacturer’s instructions. Sequences were read at the Gene Pool facility (University of Edinburgh) and analyzed using SSE v1.0 software. Sequences obtained in this study have been assigned the GenBank accession numbers JQ664502–JQ664509.

Phylogenetic trees were constructed using a bootstrap neighbour-joining method with 100 replications and incorporating the Kimura-2-Parameter model of nucleotide substitution and a uniform rate variation among sites using the MEGA 5.01 [35] software package with pairwise deletion for missing data. Tree construction involved two datasets of 31 complete HBV genome sequences from Pan t. troglodytes, Pan t. elioti, Pan t. verus, Pan t. schweinfurthii and Gorilla gorilla GenBank sequences and final phylogenetic analysis included representative human HBV sequences of genotypes A–H.
Recombination analysis was carried out using Tree Order Scan package of SSE v1.0 [Manuscript in preparation] generating an image of unique sequence positions in phylogenetic trees generated from sequential 250-base sequence fragments, incrementing by 25 bases. Changes to the sequence order due to changes in phylogeny at the 70% bootstrap level are reported.

Acknowledgments

The authors would like to thank the Cameroon Ministry of Forestry and Fauna, which provided the necessary permits for this work, Ape Action Africa, and the US Embassy in Cameroon for their support. N.D.W. is supported by the NIH Director’s Pioneer Award (DP1-OD000370).

Author Contributions

Conceived and designed the experiments: SL, CS, PS. Performed the experiments: SL. Analyzed the data: SL, CS. Contributed reagents/materials/analysis tools: SL, CS, MLB, CFD. Wrote the paper: SL. Animal handling and sample collection: MLB, CFD, JAK, FL, TGB, UT, JF. GVFH global site director: NDW.

References

Although the origin of hepatitis C virus infections in humans remains undetermined, a close homolog of this virus, termed canine hepacivirus (CHV) and found in respiratory secretions of dogs, provides evidence for a wider distribution of hepaciviruses in mammals. We determined frequencies of active infection among dogs and other mammals in the United Kingdom. Samples from dogs (46 respiratory, 99 plasma, 45 autopsy samples) were CHV negative by PCR. Screening of 362 samples from cats, horses, donkeys, rodents, and pigs identified 3 (2%) positive samples from 142 horses. These samples were genetically divergent from CHV and nonprimate hepaciviruses that horses were infected with during 2012 in New York state, USA. Investigation of infected horses demonstrated nonprimate hepacivirus persistence, high viral loads in plasma ($10^5$–$10^7$ RNA copies/mL), and liver function test results usually within reference ranges, although several values ranged from high normal to mildly elevated. Disease associations and host range of nonprimate hepaciviruses warrant further investigation.

Hepatitis C virus (HCV) is a positive-sense RNA virus, classified as the type member of the family Flaviviridae and the genus Hepacivirus. HCV is a major human pathogen, causing persistent infections that target and eventually destroy the liver in a substantial proportion of those infected (1,2). HCV infections are distributed worldwide and have spread epidemically within the past 40–60 years within western countries through blood-borne routes such as blood and blood product transfusion and injection drug use.

HCV shows considerable genetic diversity; 7 genotypes show >30% nt sequence divergence from each other (3). Several of these genotypes are associated with suspected endemic source areas in central and western sub-Saharan Africa (genotypes 1, 2, and 4) (4–7) and Southeast Asia (genotypes 3 and 6) (8,9). These regions harbor the greatest diversity of HCV subtypes, implying a long-term, endemic circulation of the virus for several hundred years. The spread of certain genotype variants from these populations, such as 1a and 1b to Western countries, 3a among injection drug users in Europe, and 4a to Egypt, where it was extensively transmitted by medical injections (10–14), show several parallels with the emergence and rapid spread of HIV-1 among new risk groups from a central African reservoir over a similar time frame (15).

Based on this model, it has been frequently speculated that HCV could have an ultimate animal origin in >1 nonhuman primate species, in much the same way as human HIV-1 originated from chimpanzees (15). Three of the Pan troglodytes chimpanzee subspecies are frequently infected in the wild with a lentivirus that ultimately was derived from simian immunodeficiency viruses infecting Old World monkey species (16). The hypothesis for an equivalent nonhuman primate origin for HCV fueled several published (17,18) and unpublished (S. Lyons et al., unpub. data) surveys for HCV or HCV-like viruses in chimpanzees, other apes, and a variety of Old World monkey species (19). These studies were encouraged by the serendipitous detection of a virus distantly related to HCV, termed GB virus B (GBV-B), in a laboratory-housed tamarin, a New World monkey (20,21). This detection was speculated to represent, in evolutionarily terms, the New World monkey homolog of HCV, a scenario that supports the possibility for the widespread distribution of further HCV-like viruses among Old World monkey species in Africa and Asia. Despite the plausibility of this hypothesis, we found that no survey to date has detected or obtained serologic evidence for infection with HCV or HCV-like viruses in any ape or Old World monkey species screened in published data (17,18) or in the aforementioned unpublished data of
Lyons et al. As a further puzzling observation, GBV-B infection has not been reported in any other tamarin or other New World monkey species either in the wild or among captive animals. Like HCV, its origin remains unknown.

Considering this background and previous focus on primates for HCV origins, it came as a complete surprise that a virus, much more closely related to HCV than GBV-B, was recently described for domestic dogs (22). Its host and apparent tropism for the respiratory tract (and potential association with infectious respiratory disease) represent major differences from what might be expected for a close relative of HCV. It came as a further surprise that CHV infections were detected in plasma samples from 8 of 103 horses in New York state, USA (23). Results of a novel serologic test of samples from horses for antibodies to the conserved nonstructural protein 3 (NS3) showed an overall seropositivity of 35%, and 80 samples from dogs were negative by serologic testing and PCR. The wider host range of the virus implied by these findings led the authors to propose a new name, non-primate hepacivirus (NPHV) for CHV, and this new nomenclature is used in the current study.

To investigate the species distribution of NPHV or homologs in a range of mammalian species and to investigate clinical features of infection, we initiated large-scale PCR-based screening of plasma, respiratory, and postmortem liver, spleen, and lung samples from horses, dogs, cats, and other species originating in the United Kingdom. The PCR was specific for conserved sequences in the 5′ untranslated region (5′-UTR) and the NS3 regions of NPHV. NPHV was detected in 3 horses (horse 1, horse 2, and horse 3); initial variability observed in the published NPHV sequences was indicative of NS3 antigen used in a previous study (23). To validate the PCR, RNA transcripts were generated from a plasmid containing partial CHV NS3 cDNA by using the Ambion T7 transcription kit (Promega Corp., Southampton, UK). Transcripts were purified with the RNeasy kit (QIAGEN), and concentrations were determined by using the NanoDrop 2000 (NanoDrop Products, Wilmington, DE, USA). RNA extractions were performed on 140 μL of plasma or respiratory sample by using the QIAamp viral extraction kit (QIAGEN) according to the manufacturer’s instructions and eluted in a final volume of 60 μL. All tissue samples were homogenized in lysis buffer; RNA was extracted by using the RNeasy Mini Kit (QIAGEN) according to instructions and eluted in a final volume of 60 μL. Peripheral blood mononuclear cells were separated from whole blood immediately after collection by centrifugation on a Ficoll-Hypaque density gradient by using Histopaque 1077 according to manufacturer’s instructions (Sigma Aldrich, St. Louis, MO, USA), and RNA was extracted by using QIAamp RNA blood mini kit as instructed (QIAGEN) and eluted in final volume of 100 μL.

DNA was converted to cDNA by using random hexamers with the Reverse Transcription System A3500 (Promega) and then used in nested PCR with previously published CHV NS3 primers (22) Chv-0F1,Chv-0R1S1, Chv-0F2, and Chv-0R2 and new equine-based NS3 primers (Table 1) and amplified by using 2 rounds of 30 cycles at 94°C for 18 s, 50°C for 21 s, and 72°C for 1.5 min; and 1 cycle of 72°C for 5 min, with 2 μL of first-round amplicon added to the second round. Degenerate equine- and canine-based NS3 primers were designed on the basis of the sequence variability observed in the published NPHV sequences and used to additionally screen all samples from dogs and equids. The CHV NS3 transcript was tested by using both NS3 primer sets and used as a control in screening, with sensitivity of 0.5–5 RNA copies in a reaction (Tables 1,2).
concentrations of 10 lutions were prepared from the CHV NS3 transcript from generated from a dilution series of the NS3 transcript. Direct-time quantitative PCR and a standard calibration curve obtained in this study have been assigned the GenBank analyzed by using SSE version 1.0 software (to the manufacturer’s instructions. Sequences were executed by using Big Dye Terminator sense primers used in the second round of amplification in both directions by using the inner sense and inner antisense primer sets: EQ5
2 uL was added to the second-round PCR with respect to the 5,000 2/2 2/2 500 2/2 2/2 50 2/2 2/2 5 2/2 2/2 0.5 0/2 0/2 0 0/2 0/2

*NS, nonstructural protein.

To confirm positive results of screening, we designed degenerate primers derived from the NPHV sequences for the 5'-UTR and NS5B (Table 1). For all detected positive results, we used SuperScript III One-Step RT-PCR (Life Technologies, Paisley, UK) with 6 µL of RNA and cycling conditions as published (23) with 1 of the following first-round primer sets: EQ5→UTRAS and EQ5→UTROAS or EQNS5BIS and EQNS5BIAS. From the first round, 2 µL was added to the second-round PCR with respective forward and reverse primer sets: EQ5→UTRIS and EQ5→UTRIS2 or EQNSBIS2 and NS5BIAS2, with the following cycling conditions: 30 cycles at 94°C for 18 s, 50°C for 21 s, and 72°C for 1.5 min; and 1 cycle of 72°C for 5 min.

Positive second-round PCR amplicons were sequenced in both directions by using the inner sense and inner antisense primers used in the second round of amplification. Sequencing was executed by using Big Dye Terminator version 3.1 (Applied Biosystems, Paisley, UK) according to the manufacturer’s instructions. Sequences were analyzed by using SSE version 1.0 software (24). Sequences obtained in this study have been assigned the GenBank accession nos. JX948116–JX948121.

Viral loads of positive samples were determined by real-time quantitative PCR and a standard calibration curve generated from a dilution series of the NS3 transcript. Dilutions were prepared from the CHV NS3 transcript from concentrations of 10^6 to 1 copy/µL; 5 µL of transcript RNA was used to generate cDNA by using random hexamers and reverse transcription. Five-microliter aliquots of cDNA were assayed in triplicate for the 3 positive samples in the same way. To quantify viral loads of positive samples, EQNS3IS and EQNS3IAS primers were used with 4 µL of cDNA in the SensiFAST SYBR Hi-ROX Kit (Bio-Line, London, UK) per manufacturer’s instructions (Table 1) with the exception that the annealing temperature was reduced to 50°C and the extension time extended to 15 s. Samples were analyzed in triplicate and fluorescence measured by using the Rotor-Gene Q system (QIAGEN). Viral loads were read from the standard curve generated and converted to RNA copies/mL for sample volume used in extraction and elution of the RNA.

Results

Sample Screening

To investigate the frequency of NPHV infection in dogs, 46 respiratory samples collected from dogs over a 6-month period in the Edinburgh area were screened by published PCR-based screening methods (22,23) by using primers from the 5'-UTR and NS3 regions. An RNA transcript from the NS3 region (23) verified the sensitivity of the NS3-based assay to single RNA copies per amplification reaction (Table 2). All samples were negative in both genome regions (Table 3). Ninety-nine plasma samples from dogs that had a variety of clinical conditions and had been referred for virology screening, along with 15 autopsy lung, liver, and spleen samples from dogs, were additionally screened; results were uniformly negative in both regions.

Since publication of the NS3- and 5'-UTR–based PCRs (22), comparative sequence data from several NPHV-infected horses have become available (23). These data revealed sequence variability in the primer binding regions of both primer sets. We therefore redesigned the screening primers in both regions (Table 1) to accommodate this. In the 5'-UTR region, it was additionally possible to ensure that primers matched homologous regions of HCV genotypes 1–7. The new nested NS3 primers showed similar sensitivity for the NS3 transcript (Table 2). These new

1978 Emerging Infectious Diseases • www.cdc.gov/eid • Vol. 18, No. 12, December 2012

Table 1. Transcript titration of nonprimate hepaciviruses with NS3 primers in samples from domestic horses, United Kingdom

<table>
<thead>
<tr>
<th>Transcript RNA copies/mL</th>
<th>Published NS3</th>
<th>New NS3</th>
</tr>
</thead>
<tbody>
<tr>
<td>5 x 10^4</td>
<td>2/2</td>
<td>2/2</td>
</tr>
<tr>
<td>5 x 10^5</td>
<td>2/2</td>
<td>2/2</td>
</tr>
<tr>
<td>5 x 10^6</td>
<td>2/2</td>
<td>2/2</td>
</tr>
<tr>
<td>5,000</td>
<td>2/2</td>
<td>2/2</td>
</tr>
<tr>
<td>500</td>
<td>2/2</td>
<td>2/2</td>
</tr>
<tr>
<td>50</td>
<td>2/2</td>
<td>2/2</td>
</tr>
<tr>
<td>5</td>
<td>2/2</td>
<td>2/2</td>
</tr>
<tr>
<td>0.5</td>
<td>0/2</td>
<td>0/2</td>
</tr>
<tr>
<td>0</td>
<td>0/2</td>
<td>0/2</td>
</tr>
</tbody>
</table>

NS, nonstructural protein.

Table 2. Nonprimate hepaciviruses primer sequences for 5'-UTR, NS3, and NS5B in samples from horses, United Kingdom

<table>
<thead>
<tr>
<th>Primer</th>
<th>Position</th>
<th>Sequence, 5’→3’</th>
</tr>
</thead>
<tbody>
<tr>
<td>EQ5→UTROS</td>
<td>Forward outer sense</td>
<td>ACA YYA CCA TGT GTC ACT CCC CCT</td>
</tr>
<tr>
<td>EQ5→UTROAS</td>
<td>Reverse outer antisense</td>
<td>CYC ATG TCC TAT GGT CTA CGA GA</td>
</tr>
<tr>
<td>EQ5→UTRIS</td>
<td>Forward inner sense</td>
<td>ACA CGG AAA YGG GTT AAC CAY ACY C</td>
</tr>
<tr>
<td>EQ5→UTRIAS</td>
<td>Reverse inner antisense</td>
<td>GCC CTC GCA AGC ATC TCA TCA G</td>
</tr>
<tr>
<td>EQNS3OS</td>
<td>Forward outer sense</td>
<td>ATW TGT GAT GAR TGC CAY AGY AC</td>
</tr>
<tr>
<td>EQNS3OAS</td>
<td>Reverse outer antisense</td>
<td>TAG TAG GTB ACA GCR TTA GGY CC</td>
</tr>
<tr>
<td>EQNS3IS</td>
<td>Forward inner sense</td>
<td>TGY AAR GTG GTD AAG CTT GTT GT</td>
</tr>
<tr>
<td>EQNS3IAS</td>
<td>Reverse inner antisense</td>
<td>TGG CAG AAG YTA AGR TGY CTY CC</td>
</tr>
<tr>
<td>EQNS5BIS</td>
<td>Forward outer sense</td>
<td>AAR TGY TTT GAC TGY ADB GTC ACT C</td>
</tr>
<tr>
<td>EQNS5BIS2</td>
<td>Reverse inner antisense</td>
<td>ACT RTG ACT RAT YGT YTC CCA ACT CG</td>
</tr>
<tr>
<td>EQNS5BIAS2</td>
<td>Forward inner sense</td>
<td>CAY GAT ATA GAH ACT GAG AGR GA</td>
</tr>
<tr>
<td>EQNS5BIAS2</td>
<td>Reverse inner antisense</td>
<td>TCR TCT TCC TCR ACG CCY TTR CTG G</td>
</tr>
</tbody>
</table>

*UTR, untranslated region; NS, nonstructural protein.
Hepaciviruses in horses

Primer sets were used to repeat screening of all canine respiratory, plasma, and autopsy samples that produced uniformly negative results (Table 3).

To investigate the possible infection of non-canine mammalian species, we screened available plasma/serum, respiratory, and liver samples from horses (n = 175), donkeys (n = 16), domestic cats (n = 56), pigs (n = 40), and wild mice (n = 61) by using both sets of conserved primers (Table 3). From this extended survey, 3 plasma samples from horses were positive on initial screening and confirmed positive on reextraction and reamplification in 5’-UTR and NS3 regions. PCR of samples of all types from all other studied mammalian species showed negative results.

To confirm the presence of NPHV sequences in the 3 screen-positive horses, we further amplified each sample using conserved primers in the NSSB region and comparing amplified sequences from each region with homologous regions of previously identified positive horses (Figure). Although this method does not represent a comprehensive genetic analysis, these sequence comparisons demonstrated that each of the positive horses was infected with NPHV variants distinct from the transcript-positive control and from each of the 8 previously identified infected horses in the USA. All 3 variants showed similar branching orders in each genome region, consistent with the observed lack of recombination in previous analyses (23).

Virologic and Clinical Examination of NPHV-positive Horses

The 3 infected horses originated in Scotland and comprised 2 geldings and 1 mare, 12–20 years of age (Table 4). Clinical records of each horse from the time of sample collection failed to identify evidence of hepatitis or systemic disease. Liver function tests provided no evidence for hepatic inflammation: γ-glutamyl transferase (GGT) and glutamate dehydrogenase were within reference range, except a mildly elevated GGT level in horse 2 (Table 5). Testing also ruled out hepatic insufficiency: bile acid lev-

Table 3. Nonprimate hepaciviruses sequences detected by using PCR on mammal samples, United Kingdom.

<table>
<thead>
<tr>
<th>Animal/sample type</th>
<th>No.</th>
<th>NS3 5’-UTR</th>
<th>Published NS3</th>
<th>New primers, 5’-UTR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dog</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Respiratory</td>
<td>53</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Plasma</td>
<td>99</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Lung</td>
<td>15</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Liver</td>
<td>15</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Spleen</td>
<td>15</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Horse</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Respiratory</td>
<td>40</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Plasma</td>
<td>142</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Donkey</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Plasma</td>
<td>16</td>
<td>NA</td>
<td>NA</td>
<td>0</td>
</tr>
<tr>
<td>Cat</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Plasma</td>
<td>56</td>
<td>NA</td>
<td>NA</td>
<td>0</td>
</tr>
<tr>
<td>Pig</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Serum</td>
<td>40</td>
<td>NA</td>
<td>NA</td>
<td>0</td>
</tr>
<tr>
<td>Mouse</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Liver</td>
<td>61</td>
<td>NA</td>
<td>NA</td>
<td>0</td>
</tr>
</tbody>
</table>

*Data from 23. NS, nonstructural protein; UTR, untranslated region; NA, not applicable.

Figure. Phylogenetic analysis of A) 5’ untranslated region, B) nonstructural protein 3, and C) nonstructural protein 5B regions of nonprimate hepatitis virus sequence amplified from screen-positive study animals. Neighbor-joining trees of nucleotide sequences from each genome region were constructed from Jukes-Cantor corrected pairwise distances calculated by using the program MEGA version 5 (25; datasets were bootstrap re-sampled 500× to indicate robustness of branching (values ≥70% shown on branches). The hepacivirus genotype 1a sequence, M62321, was used to root the tree (not shown). Scale bars indicate nucleotide substitutions per site.
Table 4. Clinical features of domestic horses infected by nonprimate hepaciviruses, United Kingdom*

<table>
<thead>
<tr>
<th>Horse</th>
<th>Sample</th>
<th>Collection date</th>
<th>Area</th>
<th>Age, y/sex</th>
<th>Presenting disease</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>EF_317/98</td>
<td>1998</td>
<td>Caithness</td>
<td>8/F</td>
<td>Lameness</td>
</tr>
<tr>
<td>2</td>
<td>EF_330/97</td>
<td>1997</td>
<td>Perthshire</td>
<td>12/M</td>
<td>Inflammatory airway disease</td>
</tr>
<tr>
<td>3</td>
<td>EF_369/11</td>
<td>2011 Dec 2</td>
<td>Lothian</td>
<td>20/M</td>
<td>Lameness; no lung disease</td>
</tr>
<tr>
<td>4</td>
<td>EF_374/12</td>
<td>2012 Mar 1</td>
<td></td>
<td></td>
<td>Lameness; no lung disease</td>
</tr>
<tr>
<td>5</td>
<td>EF_523/12</td>
<td>2012 Mar 23</td>
<td></td>
<td></td>
<td>Lameness; no lung disease</td>
</tr>
</tbody>
</table>

*Table 5. Laboratory indices for domestic horses infected with nonprimate hepaciviruses, United Kingdom

<table>
<thead>
<tr>
<th>Horse</th>
<th>Sample</th>
<th>GGT (&lt;40 U/mL)</th>
<th>GLDH (&lt;10 U/mL)</th>
<th>Bile acids (&lt;10 U/mL)</th>
<th>Viral load, copies/mL</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>EF_317/98</td>
<td>15</td>
<td>2</td>
<td>1</td>
<td>1.3 × 10^5</td>
</tr>
<tr>
<td>2</td>
<td>EF_330/97</td>
<td>59</td>
<td>2</td>
<td>1</td>
<td>4.4 × 10^3</td>
</tr>
<tr>
<td>3</td>
<td>EF_369/11</td>
<td>15</td>
<td>2</td>
<td>6.4</td>
<td>4.8 × 10^4</td>
</tr>
<tr>
<td>4</td>
<td>EF_374/12</td>
<td>36</td>
<td>1</td>
<td>7.4</td>
<td>2.1 × 10^7</td>
</tr>
<tr>
<td>5</td>
<td>EF_523/12</td>
<td>24</td>
<td>4</td>
<td>6.3</td>
<td>7.1 × 10^4</td>
</tr>
</tbody>
</table>

*GGT, γ-glutamyl transferase; GLDH, glutamate dehydrogenase.
observed on random sampling if infections had rapidly resolved. In this respect, NPHV may at least partly reproduce the documented high rates of HCV persistence, in which >50% of those exposed showed decades- or life-long viremia and active liver disease in the absence of treatment. GBV-B, although clearly hepatotropic, does not establish persistent infections in tamarins or owl monkeys (20,27,28). However, more recent studies have demonstrated long-term persistence among experimentally infected marmosets (29,30).

Neither the clinical signs nor the liver function tests of the 3 NPHV-infected horses provided a clear indication of the organism’s association with hepatic or other systemic disease (Tables 4, 5). GGT and glutamate dehydrogenase are sensitive markers of liver inflammatory processes in the horse but with 1 exception were in the reference range. Reference levels of bile acids similarly demonstrated adequate liver function. Although the sample size was small, these relatively normal liver indices contrast with the frequent GGT and ALT elevations associated with chronic HCV infection and found in New World monkeys experimentally infected with GBV-B. Although in the current study, UK veterinary regulations did not permit liver biopsies to be performed on horses without evidence of liver disease, the current findings do not rule out a lower grade infection or potential replication in the liver without the associated immunologic response to HCV that is primarily responsible for liver damage (31,32). Future studies, perhaps refocused on NPHV screening of horses with idiopathic liver disease that have undergone biopsy sampling and have been clinically characterized, are needed to investigate further the potential for hepatotropic NPHV and manifest its clinical effects. In the longer term, and acknowledging that the horse is not the ideal experimental animal, inoculation of horses with NPHV and subsequent monitoring for viremia development, liver function abnormalities, and B- and T-cell immune responses would provide further insights into the nature of NPHV infections and associated immune response and similarity of these developments to current observations for HCV and GBV-B.

Acknowledgments

We thank the clinicians at the Royal (Dick) School of Veterinary Studies, University of Edinburgh, for their assistance in sourcing a wide variety of samples.

Global Viral Forecasting in this study was supported by the US Department of Defense Armed Forces Health Surveillance Center, Division of Global Emerging Infections, Surveillance Operations and the Defense Threat Reduction Agency Cooperative Biological Engagement Program (DTRA-CBEP), Google.org, the Skoll Foundation, and the U.S. Agency for International Development (USAID) Emerging Pandemic Threats Program, PREDICT project, under the terms of Cooperative Agreement no. GHN-A-00-09-00010-00.

Ms Lyons is completing a PhD degree in virology in the Royal (Dick) School of Veterinary Studies at University of Edinburgh. Her research interests are focused on the evolutionary history, cross-species transmission, and host adaptation of human viruses and their primate/nonprimate homologs; including hepatitis B, GB, and hepatitis C viruses.

References


Address for correspondence: Peter Simmonds, Infection and Immunity Division, Roslin Institute, University of Edinburgh, Easter Bush, Edinburgh, EH25 9RG, UK; email: peter.simmonds@ed.ac.uk
Viraemic frequencies and seroprevalence of non-primate hepacivirus and equine pegiviruses in horses and other mammalian species

Sinéad Lyons,1 Amit Kapoor,2 Bradley S. Schneider,3 Nathan D. Wolfe,3 Geoff Culshaw,1 Brendan Corcoran,1 Andy E. Durham,4 Faith Burden,5 Bruce C. McGorum1† and Peter Simmonds1†

†These authors contributed equally to this work.

1Roslin Institute, University of Edinburgh, Easter Bush, Edinburgh, UK
2Center for Infection and Immunity, Mailman School of Public Health, Columbia University, New York, NY 10032, USA
3Metabiota, 1 Sutter Street, Suite 600, San Francisco, CA 94104, USA
4Liphook Equine Hospital, Forest Mere, Liphook, Hampshire, UK
5The Donkey Sanctuary, Sidmouth, Devon, UK

Non-primate hepacivirus (NPHV), equine pegivirus (EPgV) and Theiler’s disease associated virus (TDAV) are newly discovered members of two genera in the Flaviviridae family, Hepacivirus and Pegivirus respectively, that include human hepatitis C virus (HCV) and human pegivirus (HPgV). To investigate their epidemiology, persistence and clinical features of infection, large cohorts of horses and other mammalian species were screened for NPHV, EPgV and TDAV viraemia and for past exposure through serological assays for NPHV and EPgV-specific antibodies. NPHV antibodies were detected in 43% of 328 horses screened for antibodies to NS3 and core antibodies, of which three were viraemic by PCR. All five horses that were stablemates of a viraemic horse were seropositive, as was a dog on the same farm. With this single exception, all other species were negative for NPHV antibodies and viraemia: donkeys (n = 100), dogs (n = 112), cats (n = 131), non-human primates (n = 164) and humans (n = 362). EPgV antibodies to NS3 were detected in 66.5% of horses, including 10 of the 12 horses that had EPgV viraemia. All horse and donkey samples were negative for TDAV RNA. By comparing viraemia frequencies in horses with and without liver disease, no evidence was obtained that supported an association between active NPHV and EPgV infections with hepatopathy. The study demonstrates that NPHV and EPgV infections are widespread and enzootic in the study horse population and confirms that NPHV and potentially EPgV have higher frequencies of viral clearance than HCV and HPgV infections in humans.

INTRODUCTION

Hepatitis C virus (HCV) was discovered over 20 years ago in the serum of a patient with non-A, non-B hepatitis. HCV is a major cause of chronic liver disease, hepatocellular carcinoma and liver cirrhosis, infecting approximately 3% of the world’s population (Alter et al., 1999; Perz et al., 2004; Shepard et al., 2005). Human pegivirus (HPgV) is a lymphotropic virus that causes little or no apparent disease in humans despite persistently infecting an estimated 5% of the human population (Berg et al., 1999).

HCV and HPgV are positive-stranded RNA viruses with genomes approximately 9000 bases in length which are classified into two separate genera of the family Flaviviridae, Hepacivirus and Pegivirus respectively (King et al., 2011; Stapleton et al., 2011). Until recently, the only other virus classified as a member of the genus Hepacivirus was GBV-B, detected in a laboratory tamarin (Muerhoff et al., 1995; Simons et al., 1995a) in which it caused hepatitis and could be experimentally transmitted to certain other New World primate species (Bukh et al., 2001). GBV-B has, however, never been detected in any other captive or wild primate populations, in South America or elsewhere (Simmonds,
By contrast, pegiviruses infect a diverse range of mammals including humans, New World monkeys, chimpanzees (SPgV) and bats (BPgV) (Adams et al., 1998; Birkenmeyer et al., 1998; Epstein et al., 2010; Kapoor et al., 2013a; Quan et al., 2013; Simons et al., 1995b).

Very recently, several studies of dogs, horses, wild rodents, bats and non-human primates (NHPs) have revealed much greater viral diversity of members of the genera Hepacivirus and Pegivirus (Burbelo et al., 2012; Drexler et al., 2013; Kapoor et al., 2011, 2013a; Lauck et al., 2013; Lyons et al., 2012; Quan et al., 2013). Homologues of HCV were first detected in domestic dogs and horses, described as non-primate hepaciviruses (Burbelo et al., 2012; Lyons et al., 2012), subsequently in several species of bats (bat hepacivirus, BHV; Quan et al., 2013) and rodents (rodent hepacivirus, RHV; Drexler et al., 2013; Kapoor et al., 2013b), and most recently in the black-and-white colobus monkey (Lauck et al., 2013). While clearly similar in genome organization to HCV and containing homologues of each of the structural and non-structural genes characterized in HCV, these new animal viruses are remarkably divergent in sequence from both HCV and GBV-B, and are likely to qualify as putative additional species within the genus Hepacivirus. At least two lineages of rodent hepaciviruses additionally contain an internal ribosomal entry site (IRES) that more closely resembles those of pegiviruses than the type IV IRESs found in all other hepaciviruses (Drexler et al., 2013).

Similarly, several new pegiviruses have been described in horses (equine pegivirus, EPgV; Kapoor et al., 2013a) several species of rodents (RPgV; Drexler et al., 2013; Kapoor et al., 2013b) and further species of bats (Quan et al., 2013). In contrast to the close host associations of primate and horse pegiviruses, rodent viruses were highly genetically heterogeneous and lineages were interspersed with viruses infecting other mammalian species, providing evidence against the previous hypothesis for virus–host co-evolution (Sharp & Simmonds, 2011).

Despite the differences in pathogenicity between HCV and HPgV infections, both are capable of establishing persistent infections in humans, as can non-primate hepacivirus (NPHV) in horses, with persistent viraemia detected in a horse serially sampled over 6 months (Lyons et al., 2013). The virus, termed Theiler’s disease associated virus (TDAV), was proposed as the aetiological agent of acute hepatitis observed in horses following administration of equine plasma (Chandriani et al., 2013). TDAV was detected in three separate outbreaks of acute serum hepatitis in horses in the USA, and subsequent inoculation of horses with TDAV-positive plasma resulted in several weeks of viraemia preceding liver disease.

In the current study we investigated frequencies of active and past infection of horses and other mammalian species by NPHV, and extended screening for EPgV and TDAV to horses and donkeys. We also made a preliminary investigation of the potential association of NPHV and EPgV infections and hepatopathy in horses.

RESULTS

NPHV serology screening

To investigate frequencies of current infection and past exposure to NPHV, we developed an ELISA using recombinant proteins expressed from the NS3-helicase domain and from the core gene (Table S3 available in the online Supplementary Material). This two-stage assay was used to screen a large study group of horses and other mammalian species (328 horses, 100 donkeys, 116 dogs, 132 cats, 362 human, 164 NHPs; Table 1). All samples were screened in blocking concentrations of soluble non-recombinant pGex-2t Escherichia coli lysate, and reactivity to NS3/core-expressing and non-expressing controls was compared with minimize assay non-specificity. The cut-off for the NPHV assay was set conservatively as the mean serological reactivity (measured at 405nm) plus three sds of unreactive samples (NS3: 0.128 + 3(0.053) = 0.287; core: 0.136 + 3(0.048) = 0.280), and was used to categorize samples as anti-NPHV positive or negative.

All samples were screened for both anti-NS3 and anti-core IgG antibodies. Samples were considered positive if above the cut-off in both ELISAs; those that were solely NS3

<table>
<thead>
<tr>
<th>Species</th>
<th>NPHV PCR+</th>
<th>Ab+</th>
<th>EPgV PCR+</th>
<th>Ab+</th>
<th>TDAV PCR+</th>
</tr>
</thead>
<tbody>
<tr>
<td>Horses</td>
<td>0/111</td>
<td>53/111</td>
<td>7/111</td>
<td>62/111</td>
<td>0</td>
</tr>
<tr>
<td>Hepatopathy</td>
<td>3/217</td>
<td>89/217</td>
<td>5/217</td>
<td>130/217</td>
<td>0</td>
</tr>
<tr>
<td>Control</td>
<td>0/100</td>
<td>0/100</td>
<td>0/100</td>
<td>0/100</td>
<td>NA</td>
</tr>
<tr>
<td>Donkeys</td>
<td>0/362</td>
<td>0/362</td>
<td></td>
<td></td>
<td>NA</td>
</tr>
<tr>
<td>Humans</td>
<td>0/164</td>
<td>0/164</td>
<td></td>
<td></td>
<td>NA</td>
</tr>
<tr>
<td>Non-human primates</td>
<td>0/113</td>
<td>1/113</td>
<td></td>
<td></td>
<td>NA</td>
</tr>
<tr>
<td>Dogs</td>
<td>0/131</td>
<td>0/131</td>
<td></td>
<td></td>
<td>NA</td>
</tr>
<tr>
<td>Cats</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>NA</td>
</tr>
</tbody>
</table>

+, Positive; NA, not applicable; Ab, antibody.
antibody positive or core antibody positive were classified as indeterminate (Table 2). A plot of the serology results for the NPHV NS3 absorbance at A405nm (A) against NPHV core A absorbance at A405nm (Fig. 1) demonstrated good concordance between reactivity to NS3 and core antigens (R²=0.5285). Overall, 142 from 327 horse samples tested in both assays were confirmed positive (43.3%) and 19 (5.5%) were reactive in only one assay. In view of their indeterminate status, the latter 19 samples were excluded from subsequent analyses of disease status. All samples from horses previously identified as viraemic were reactive with both antigens by ELISA (Lyons et al., 2012; Table 3).

All samples from all other mammalian species were unreactive in both NS3 and core ELISAs with the exception of a single sample taken from a dog that had come into regular contact with the previously reported NPHV viraemic horse (Table 3; A: NS3, 1.003 and core, 0.6255) (Lyons et al., 2012). Serial samples collected from a persistently viraemic horse showed declining antibody levels over a 1-year period (horse 3; Figs 1, 2 and Table 3). Over the initial 6-month period, viral loads were elevated but declining as previously reported (Lyons et al., 2012). At the time of final sampling, RNA was undetectable indicating viral clearance, but the horse remained seropositive. Similarly, all five contact horses co-housed with the viraemic horse were non-viraemic but seropositive (Table 3).

Infection frequencies of EPgV in horses and donkeys.

An assay to detect antibodies to EPgV was developed by expression of the region of NS3 in E. coli (pET-28b) homologous to that used for NPHV antibody screening. Recombinant protein was used in an indirect ELISA format using the same blocking and control antigens used for NPHV screening to minimize assay non-specificity. Due to the lack of a comparable core region for EPgV, Western blots (WB) using recombinant NS3 protein were used to confirm the presence of antibody in samples reactive to the EPgV-NS3 ELISA. Samples from 328 horses and 100 donkeys were initially tested (Table 1). Using an A threshold of the mean serological reactivity (0.130) + 3 SD of unreactive samples (0.262), 63% of samples from horses but none of the 100 donkey samples were identified as reactive by ELISA (Fig. 3). Screen-positive samples were tested for anti-NS3 antibodies by WB, of which 88% were confirmed and considered positive (58.5%). NS3 ELISA-reactive samples that were negative in the supplementary WB assay were considered as indeterminate and not included in analyses of disease associations.

To determine frequencies of active EPgV infection and to investigate the relationship between viraemia and presence of EPgV-specific antibodies, all samples were screened by PCR using EPgV-specific primers targeting the NS3 region (Table S1; Kapoor et al., 2013a). A transcript control based on the NS3 helicase was generated from a high titre sample and verified the sensitivity of the NS3 based PCR assay to 5 RNA copies/reaction (Table S4). Viraemia was detected in 12/328 horses (Table 1). Further PCRs based on the published 5’ UTR and NS5B of EPgV confirmed the presence of viraemia in all positive samples (Table S1). All PCR products were sequenced and their genetic relatedness compared with published sequences of EPgV and the recently isolated rodent pegivirus and bat pegivirus (Kapoor et al., 2013b; Quan et al., 2013). All EPgV isolates were distinct from those isolated in the USA and were also distinct from the single available TDAV sequence (Fig. 4a, b). All but two of the RNA positive horses were seropositive by ELISA and WB (Fig. 3).

The 12 EPgV viraemic horses comprised five horses (five females, age 5–9 years) from a single site in Scotland and

### Table 2.
(a) Confirmation of serology-reactive samples by confirmatory ELISAs and Western blot. (b) NPHV and EPgV RNA detection in serology-positive horses

<table>
<thead>
<tr>
<th>NPHV</th>
<th>EPgV</th>
</tr>
</thead>
<tbody>
<tr>
<td>NS3 Ab+</td>
<td>NS3 Ab-</td>
</tr>
<tr>
<td>Core Ab+</td>
<td>142</td>
</tr>
<tr>
<td>Core Ab-</td>
<td>3</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>NPHV</th>
<th>EPgV</th>
</tr>
</thead>
<tbody>
<tr>
<td>PCR+</td>
<td>PCR-</td>
</tr>
<tr>
<td>Ab+</td>
<td>3</td>
</tr>
<tr>
<td>Ab-</td>
<td>0</td>
</tr>
</tbody>
</table>

Ab, antibody.

### Table 3. Serological reactivity in horses with NPHV viraemia detected by PCR

<table>
<thead>
<tr>
<th>Horse</th>
<th>Sample date</th>
<th>Viral load*</th>
<th>NS3 NPHV</th>
<th>Core NPHV</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>A405nm</td>
<td>A405nm</td>
</tr>
<tr>
<td>1</td>
<td>1998</td>
<td>1.3×10^5</td>
<td>0.65</td>
<td>0.75</td>
</tr>
<tr>
<td>2</td>
<td>1997</td>
<td>4.4×10^5</td>
<td>0.53</td>
<td>0.66</td>
</tr>
<tr>
<td>3a</td>
<td>Dec 2011</td>
<td>4.8×10^2</td>
<td>0.90</td>
<td>0.69</td>
</tr>
<tr>
<td>3b</td>
<td>Mar 2012</td>
<td>2.1×10^5</td>
<td>0.67</td>
<td>0.68</td>
</tr>
<tr>
<td>3c</td>
<td>Mar 2012</td>
<td>7.1×10^2</td>
<td>0.51</td>
<td>0.62</td>
</tr>
<tr>
<td>3d</td>
<td>Dec 2012</td>
<td>Negative</td>
<td>0.50</td>
<td>0.41</td>
</tr>
<tr>
<td>3e</td>
<td>Apr 2013</td>
<td>Negative</td>
<td>0.37</td>
<td>0.38</td>
</tr>
<tr>
<td>4f</td>
<td>July 2013</td>
<td>Negative</td>
<td>1.00</td>
<td>0.63</td>
</tr>
<tr>
<td>5f</td>
<td>Mar 2012</td>
<td>Negative</td>
<td>0.61</td>
<td>0.50</td>
</tr>
<tr>
<td>6§</td>
<td>Mar 2012</td>
<td>Negative</td>
<td>0.76</td>
<td>0.70</td>
</tr>
<tr>
<td>7§</td>
<td>Mar 2012</td>
<td>Negative</td>
<td>0.79</td>
<td>0.66</td>
</tr>
<tr>
<td>8§</td>
<td>Mar 2012</td>
<td>Negative</td>
<td>0.77</td>
<td>0.59</td>
</tr>
<tr>
<td>9§</td>
<td>Mar 2012</td>
<td>Negative</td>
<td>0.70</td>
<td>0.69</td>
</tr>
</tbody>
</table>

*RNA copies ml⁻¹ determined by qRT-PCR.
†All samples labelled ‘3’ refer to serial samples from a previously identified NPHV viraemic horse taken between December 2011 and April 2013 (Lyons et al., 2012).
§Dog and §horses in regular contact with Horse 3.
seven horses (four mares, two geldings, one unknown, aged 5–29 years with one of unknown age) from separate individual premises (Table 4). Seven of 12 EPgV viraemic horses were from the hepatopathy group; four of these had only serum biochemical but no clinical indication of hepatopathy, while clinical information for the other three was unavailable. Five of 12 EPgV viraemic horses were from the control horse group, being clinically healthy and having no prior indication of hepatopathy. However, subsequent testing of the viraemic control horses revealed that two of seven had serum biochemical evidence of mild hepatopathy. Thus, overall nine of 12 EPgV viraemic horses had serum biochemical evidence of hepatopathy, namely elevation in gamma-glutamyl transferase (GGT) (8/12), glutamate dehydrogenase (GLDH) (3/12) and/or bile acids (n=2; Table 4), but all six for which there was available clinical information had no clinical signs of hepatopathy or other systemic disease. Viral load was determined by qPCR against an NS3 transcript standard and ranged from 4.09 x 10^5 to 1.98 x 10^9 RNA copies ml^-1 (Table 4). Repeat sampling of five of the initially viraemic horses demonstrated persistent infections which were maintained over a 4-month period, although with declining viral loads (Table 4). There was little evidence for hepatopathy maintained over that period with liver enzyme levels largely within the normal range, with the exception of mildly elevated GGT/GLDH/bile acids in two cases (Table 4). Proportions of EPgV seropositive horses in the hepatopathy (62/111) and control (130/217) groups were not significantly different (Fischer’s exact test, P=0.55; Table 1).

**Frequency of TDAV viraemia**

Previously published primers targeting the conserved NS3 helicase and NS5B RdRp of TDAV (Chandriani et al., 2013; EVT-162/163 and EVT-180/181/83) were used to screen all 328 horses and 100 donkeys. All samples were PCR negative using the previously published primer sets. TDAV-NS3 specific primers designed as part of this study and tested against a TDAV-NS3 transcript control also failed to detect the presence of viral RNA by nested PCR in all horse and donkey samples (Table 1). It remains to be determined whether TDAV is present in equine blood products or in horses or other species outside the USA.

**DISCUSSION**

Infection with HCV and its genetically related GBV-B virus was, until recently, thought to be confined to humans and experimentally infected New World primates, respectively. The detection of canine hepacivirus (CHV) and NPHV in dogs and horses, respectively, was the first indication of a wider host range for hepacivirus infection (Burbelo et al., 2012; Kapoor et al., 2011; Lyons et al., 2012). More recently, detection and characterization of a much wider and diverse range of hepaciviruses were found in bats, rodents and an Old World primate (Drexler et al., 2013; Kapoor et al., 2013a, b; Lauck et al., 2013; Quan et al., 2013). These new discoveries have fundamentally revised our knowledge of viral diversity and host ranges of viruses in this genus, their epidemiology and pathogenesis.

Detection of NPHV infection was initially determined by PCR screening and deep sequencing (Kapoor et al., 2011; Lyons et al., 2012) and serological detection of CHV-NS3 antibodies to detect past infections (Burbelo et al., 2012). Recombinant protein expressed from the helicase domain of CHV NS3 was used as an antigen in a luciferase immunoprecipitation system (LIPS) to screen a range of species (Burbelo et al., 2012). In the current study, recombinant protein was expressed from the evolutionarily conserved NS3 helicase and core regions of NPHV and used in

![Fig. 1. Association between serological reactivities to NS3 and core antigens used for ELISA screening in horse samples. Samples in red indicate previously reported viraemic horses. Pink indicates seroreactivity over a period of 1.5 years of serial sampling of a single viraemic case. The black line indicates the overall declining trend in the combined NS3 and core antibody detection in horse 3 (NPHV, RNA positive; Table 3). The green symbol indicates the seropositive sample from a dog. Serological reactivity was measured at A405nm.](image)
Viraemic and widely circulates in horse populations in the USA, comparable to the 35 % previously reported in the USA. NPHV antibodies were detected in 43 % of horses, which is lower than the expected seroprevalence observed in this study. The potential occurrence of re-reversion (loss of detectable antibodies over time) means that they eventually become undetectable, indicating that the NPHV antibodies declined over time in horse 3 (Table 3) and may ultimately become undetectable. The low frequency of viraemia in seropositive horses (3/142; 2.1 %) also contrasts with much higher rates of HCV persistence in humans, where over 50 % of anti-HCV seropositive individuals are actively infected (Crofts & Aitken, 1997; Nalpas et al., 1992). Through serial sampling of one study horse (horse 3; Table 3), we were able to document declining viral loads and clearance of RNA over a 1-year period from an initially viraemic horse, while remaining seropositive. Furthermore, the observation that NPHV antibodies declined over time in horse 3 (Table 3) and may ultimately become undetectable indicates that the NPHV seroprevalence observed in this study may underestimate the level of prior NPHV infections in the equine population. This limitation of serology tests to estimate past exposure.

Table 4. Viraemic frequency and seroprevalence of NPHV and EPgV.

Horses 1–5 are in the control group and horses 6–12 are in the hepatopathy group. Horses 1–5 were sampled on three occasions.

<table>
<thead>
<tr>
<th>Horse</th>
<th>Viral load*</th>
<th>EPgV Ab A405nm</th>
<th>Western blot</th>
<th>GGT† (&lt;42 U l⁻¹)</th>
<th>GLDH† (&lt;12 U l⁻¹)</th>
<th>Bile acid † (0–12 μmol l⁻¹)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.46 × 10⁶</td>
<td>0.44</td>
<td>Positive</td>
<td>0</td>
<td>4</td>
<td>10</td>
</tr>
<tr>
<td>2</td>
<td>2.79 × 10⁵</td>
<td>1.39</td>
<td>Positive</td>
<td>2</td>
<td>4</td>
<td>9</td>
</tr>
<tr>
<td>3</td>
<td>1.98 × 10⁵</td>
<td>0.97</td>
<td>Positive</td>
<td>23</td>
<td>9</td>
<td>10</td>
</tr>
<tr>
<td>4</td>
<td>3.06 × 10⁵</td>
<td>0.85</td>
<td>Positive</td>
<td>21</td>
<td>4</td>
<td>6.8</td>
</tr>
<tr>
<td>5</td>
<td>1.59 × 10⁵</td>
<td>0.79</td>
<td>Positive</td>
<td>23</td>
<td>2</td>
<td>6</td>
</tr>
<tr>
<td>6</td>
<td>2.25 × 10⁴</td>
<td>0.89</td>
<td>Positive</td>
<td>34</td>
<td>9</td>
<td>10.2</td>
</tr>
<tr>
<td>7</td>
<td>4.09 × 10⁴</td>
<td>1.62</td>
<td>Positive</td>
<td>34</td>
<td>8</td>
<td>10</td>
</tr>
<tr>
<td>8</td>
<td>1.26 × 10⁴</td>
<td>0.30</td>
<td>Positive</td>
<td>63</td>
<td>8</td>
<td>6</td>
</tr>
<tr>
<td>9</td>
<td>NA</td>
<td>NA</td>
<td>Negative</td>
<td>156</td>
<td>50</td>
<td>28</td>
</tr>
<tr>
<td>10</td>
<td>NA</td>
<td>0.55</td>
<td>Positive</td>
<td>749</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>11</td>
<td>NA</td>
<td>0.45</td>
<td>Positive</td>
<td>243</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>12</td>
<td>NA</td>
<td>0.42</td>
<td>Positive</td>
<td>120</td>
<td>NA</td>
<td>NA</td>
</tr>
</tbody>
</table>

*Viral loads were determined by qRT-PCR.
†Elevated values are shown in bold.
‡Viraemic horses which were concluded as being antibody negative (Horse 7: ELISA and WB negative) or indeterminate (Horse 8: WB negative).
NA, Not applicable.

ELISAs to determine seroprevalence of NPHV in horses, donkeys, cats, dogs, NHPs and humans. Using two separate ELISAs to detect NPHV antibodies, and considering only samples reactive in both assays to be confirmed as seropositive, substantially improved the likely specificity of the assays compared with those based on single antigens. We additionally set assay thresholds conservatively to further prevent misclassification of samples. The weak seropositivity reported in a single cow by the LIPS method (Burbelo et al., 2012) could not be confirmed by a PCR assay. A second serological assay was developed to estimate seroprevalence to EPgV by expression of the NS3 helicase; in the absence of a homologous core protein, all screen-positive samples were confirmed by WB to minimize non-specificity. Seroprevalence data provide evidence for past virus exposure, although as discussed below, the potential occurrence of re-reversion (loss of detectable antibodies over time) means that they represent minimum estimates for frequencies of past infection in study populations.

NPHV antibodies were detected in 43 % of horses, which is comparable to the 35 % previously reported in the USA (Burbelo et al., 2012). These findings indicate NPHV is enzootic and widely circulates in horse populations in the UK, France and the USA. These exposure frequencies are substantially higher than for HCV in human populations, even in regions of past epidemic spread such as in Pakistan and Egypt, where maximum seroprevalences of approximately 4 % and 15 %, respectively, have been described (Averhoff et al., 2012; Miller & Abu-Raddad, 2010; Qureshi et al., 2010). These suggest likely differences in transmission efficiency between these two hepacivirus species and are indicative of non-parenteral routes of NPHV transmission between horses. The low frequency of viraemia in seropositive horses (3/142; 2.1 %) also contrasts with much higher rates of HCV persistence in humans, where over 50 % of anti-HCV seropositive individuals are actively infected (Crofts & Aitken, 1997; Nalpas et al., 1992; Sugitani et al., 1992). Through serial sampling of one study horse (horse 3; Table 3), we were able to document declining viral loads and clearance of RNA over a 1-year period from an initially viraemic horse, while remaining seropositive. Furthermore, the observation that NPHV antibodies declined over time in horse 3 (Table 3) and may ultimately become undetectable indicates that the NPHV seroprevalence observed in this study may underestimate the level of prior NPHV infections in the equine population. This limitation of serology tests to estimate past exposure.

http://vir.sgmjournals.org

1705
extends to HCV. Antibodies to HCV also decline over time, sometimes to undetectable levels, several years after virus clearance (Takaki et al., 2000). NPHV antibodies were also detectable in all stablemates indicating previous infections among all horses on the farm.

All other species tested (Table 1) were seronegative with the exception of a single dog (Table 1, 3) which had direct contact with a previously reported NPHV viraemic horse (Lyons et al., 2012; Fig. 2 and Table 3). Identification of a seropositive dog in the UK provides the first evidence that NPHV may transmit to other species, and is concordant with the initial report of CHV/NPHV infections in dogs in the USA, identified in association with an outbreak of respiratory infections in separate geographically isolated dog kennels (Kapoor et al., 2011).

It remains to be determined conclusively how NPHV infections were transmitted between the farm horses. Virus transmission between these animals may have occurred through close contact from co-housing in the same stable, or be the result of exposure to contaminated blood in wounds or on farm surfaces, or the result of exposing the horses to contaminated needles during extensive travel abroad. Further studies on NPHV are warranted to confirm whether or not it utilizes transmission routes comparable to those of HCV; the high seroprevalences observed among horses in this and previous studies (Burvbelo et al., 2012) suggests not.

A serological assay for EPgV-specific antibodies assay was developed to determine exposure frequencies in the same cohort of horses. Anti-EPgV NS3 antibodies were confirmed by WB in 58.5 % of horses (192/328). This exposure frequency is substantially higher that the seroprevalence of HPgV in humans, globally estimated at 2–13 % in healthy blood donors (Blair et al., 1998; Gutierrez et al., 1997; Pilot-Matias et al., 1996a; Tacke et al., 1997). However, higher rates have been described in groups with frequent sexual contact, such as up to 46 % of HIV infected homosexual men (Pavlova et al., 1999; Williams et al., 2004) and 18 % among prostitutes (Scallan et al., 1998). It is necessary to note that in contrast to the anti-EPgV-NS3 assay, seroprevalence rates in humans have been measured based on anti-E2 antibodies and antibodies to the NS3 protein of HPgV have not been measured during or after
Viraemia. To date no studies have investigated pegivirus seroprevalence in bats or rodents. Active pegivirus infection was detected in 12/328 (3.8%) of horses, while all other species were EPgV RNA negative (Tables 1, 4). This detection frequency was not significantly different from the original survey of horses in the USA, where seven PCR-positive horses were found among 74 tested (9.5%; P = 0.061) (Kapoor et al., 2013a). These data are also similar to published estimates of active infection frequencies in the human population, which are approximately 5% in the developed world with upwards of 20% viraemia recorded in some developing countries (Mohr & Stapleton, 2009; Pavlova et al., 1999; Polgreen et al., 2003). Similarly, frequencies of viraemia observed in rodents and bats range from 3.9 to 5% (Kapoor et al., 2013a, 2013b; Quan et al., 2013).

Ten of the 12 RNA positive samples were seropositive for EPgV (Table 4); antibodies were not detectable by ELISA or WB for the remaining two RNA positive samples (Table 4, Fig. 3). Indeed, over the 4 month period within which horse samples were collected (Table 4), antibody levels were generally maintained despite a general declining trend.

Fig. 4. Phylogenetic analysis of (a) NS3 and (b) NS5B of EPgV sequences amplified from screen-positive study animals. Maximum-likelihood trees of nucleotide sequences were constructed using Tamura–Nei models of estimated distances calculated by using the program MEGA version 5.2. Evolutionary distances are shown on the scale bar above each tree. Datasets were bootstrap resampled 500 times to indicate robustness of branching (values > 70% shown on branches). Abbreviations detailed above correspond to; BPgV-bat pegivirus, EPgV-equine pegivirus, TDAV-Theilers disease associated virus, RPgV-rodent pegivirus, SPgV-simian pegivirus, NPHV-nonprimate hepacivirus.
in viral load. These findings contrast with reports of absent anti-E2 antibody in humans actively infected with HPgV and seroconversion on clearance of viraemia (Gomara et al., 2010; Pilot-Matias et al., 1996b; Schwarze-Zander et al., 2006; Tan et al., 1999; Van der Bij et al., 2005).

Persistent EPgV infection was evident from the repeated detection of viraemia in serial horse samples screened over a 4 month period in five horses (Table 4). Persistence of infection is characteristic of other pegiviruses, including SPgV (Simons et al., 1995b). Similarly, while immune competent individuals typically clear HPgV infections, this occurs over relatively prolonged time frames, years after primary infection (Berg et al., 1999; Tanaka et al., 1998).

The novel pegivirus TDAV was recently identified as the potential agent responsible for acute equine serum hepatitis (Chandriani et al., 2013). The USA study investigated an outbreak of acute serum hepatitis associated with the administration of equine antiserum to the botulinum toxin and found that 8/17 horses that received the antitoxin 2 preparation developed symptoms of acute hepatitis. While the screening of the study population of horses and donkeys for TDAV RNA failed to detect active infections, including in those horses with hepatitis (Table 1), no animals screened were known to have received equine blood derived products. These findings do not therefore preclude the possibility that TDAV may indeed circulate in the UK equine population, but further investigation of this may be more effectively focused on those that have received equine blood derived products.

The classification and renaming of members of the Pegivirus genus was partly in response to the accumulating failure to demonstrate any link between human infection with HPgV and hepatitis, making a description such as 'hepatitis G virus' inappropriate (Stapleton et al., 2011; Takikawa et al., 2010). Data from this study facilitated some preliminary analysis of potential associations of active NPHV and EPgV infection with hepatitis in horses. Serum biochemical evidence of hepatitis was identified in 1/3 horses with NPHV viraemia and 9/12 horses with EPgV viraemia, but further study is required to determine whether the hepatitis was attributable to NPHV and EPgV infections or to another of the many causes of equine hepatitis. The prevalence of NPHV and EPgV viraemia and seropositivity were not significantly different in horses with hepatitis versus the control group, suggesting that these viruses are unlikely to represent a common cause of hepatic disease. Further study is required to clarify potential disease associations with NPHV and EPgV infections in horses.

This study provides tools for the development of diagnostic assays for viraemia and antibody screening that will assist future research into their transmission, disease associations and tissue tropisms. The presence of hepacivirus and pegivirus species in dogs, horses, rodents and bats sheds light on the possible evolutionary history of HCV and HPgV, in which potential cross-species transmission and zoonotic origins suddenly become more plausible. Continued PCR and serology based screening of other mammalian taxa across a range of geographical locations, including areas of endemic HCV infection, is required to address these gaps in our understanding of the evolution of hepacviruses and pegiviruses.

**METHODS**

**Samples.** Serum or plasma samples (*n* = 1197) were collected from 362 humans (Cameroon), 163 NHPs (Cameroon), 328 horses (Scotland, England and France), 100 donkeys (England), 113 dogs (Scotland) and 131 cats (Scotland). Samples from humans and NHPs were collected in Cameroon by Metabiotica (formerly Global Viral Forecasting Initiative) staff in Cameroon and included 11 gorillas, 62 chimpanzees and 91 Old World monkeys as previously outlined (Lavoie et al., 2012; Lyons et al., 2012; Sharp et al., 2010). Samples from horses, donkeys, dogs and cats were archived excess diagnostic samples collected at the Royal (Dick) School of Veterinary Studies, Edinburgh, Liphook Equine Hospital Laboratory, Hampshire, and The Donkey Sanctuary, Sidmouth. Animals were of mixed breed, age and sex. Donkeys, dogs and cats had a wide range of diseases that prompted diagnostic sampling. Horses were categorized into a hepatitis group (*n* = 111) and a control group (*n* = 217). The hepatitis group comprised horses that had been investigated for suspected liver disease and which had biochemical evidence of hepatitis, as indicated by serum/plasma concentrations of gamma-glutamyltransferase (GGT), glutamate dehydrogenase (GLDH) and/or bile acids exceeding the laboratory reference upper level. Hepatopathy was attributed to a wide range of causes. The control group comprised horses for which there was no clinical suspicion of hepatopathy, including clinically healthy horses and horses that were sampled for suspected liver disease.
investigation of a wide range of diseases excluding hepatopathy. Samples were collected between 1995 and 2013, and were anonymized prior to testing.

NPHV, EpgV and TDAV viraeemia screening. All samples were screened for NPHV by PCR with previously published primers and transcript controls (Lyons et al., 2012). Horse and donkey samples were screened for EpgV and TDAV RNA. For detection of EpgV RNA, nested PCR primers were designed targeting the viral NS3 helicase based on published sequences and used in addition to published primer sets (Burbele et al., 2012). To validate the EpgV and TDAV PCR, RNA transcripts were generated from a plasmid containing PCR amplified partial NS3 cDNA by using an Ambion T7 transcription kit (Promega). Transcripts were purified and an RNase K kit (Qiagen), and concentrations were determined by using a NanoDrop 2000 (NanoDrop Products). RNA extractions were performed on 140 μl of plasma using a QiAmp viral extraction kit (Qiagen) according to the manufacturer’s instructions and eluted in a final volume of 60 μl.

RNA was converted to cDNA using random hexamers with a Superscript III Reverse Transcription System (Life Technologies) and then used in nested PCR with previously published NPHV NS3 primers (Lyons et al., 2012) in addition to newly designed EpgV NS3 and 5’UTR, and TDAV-NS3 primer sets (Table S1) and amplified using two rounds of 35 cycles at 94 °C for 18 s, 50 °C for 21 s, and 72 °C for 1.5 min; and one cycle of 72 °C for 5 min, with 2 μl of first-round amplicon added to the second round. TDAV qPCR screening was carried out as previously published (Chandriani et al., 2013). The EpgV and TDAV NS3 transcript was tested by using both EpgV and TDAV NS3 primer sets and used as a control in screening, with sensitivity of 0.5–5 RNA copies in a reaction (Tables S1, S4).

Positive second-round PCR amplicons were sequenced as previously published (Lyons et al., 2012) and analysed using SSE version 1.1 software (Simmonds, 2012).

Viral RNA quantification. Quantitative real-time PCR (qRT-PCR) was used to determine viral loads of positive samples using a standard calibration curve from a dilution series of the NS3 transcript. Dilutions of EpgV and TDAV NS3 transcript were prepared from concentrations of 10^10 to 1 copy μl⁻¹. 5 μl transcript RNA was used to generate cDNA using random hexamers and reverse transcription. Five microtitre aliquots of cDNA were assayed in triplicate for the positive samples in the same way. To quantify viral loads of positive samples, EpgV NS3S and EpgV NS3AS primers were used with 4 μl cDNA in a SensiFAST SYBR Hi-ROX kit (BioLine) as per the manufacturer’s instructions with the exception that the annealing temperature was reduced to 50 °C and the extension time extended to 15 s. Samples were analysed in triplicate and fluorescence measured by using a Rotor-Gene Q system (Qiagen). Viral loads were read from the standard curve generated and converted to RNA copies ml⁻¹ based on the sample volume used in extraction and elution of the RNA.

Synthesis of NPHV and EpgV NS3 and core recombinant proteins. Nested PCR primers for NPHV and EpgV NS3 helicase were designed based on published sequences, with appropriate restriction sites for subcloning added to the inner primers (Table S2). PCRs were performed using GoTaq (Promega) in accordance with the manufacturer’s instructions, by using the conditions outlined above for both rounds. Amplicons were cloned into a pGex-2T or pET28B vector for production of glutathione-S-transferase and polyhistidine fusion proteins respectively (Amersham Pharmacia Biotech, Merck Novagen). Briefly, the plasmid and the inserts were digested with BamHI and EcoRI according to the manufacturer’s instructions (Promega) and gel purified using a PureLink Gel Extraction kit (Invitrogen). The ligation reaction was carried out using T4 Ligase (Bioline) at an insert:vector ratio of 4 and transformed into JM109 competent cells (Agilent) for propagation. Recombinant clones were selected by antibiotic resistance and were confirmed by PCR and sequencing. High concentrations of plasmid were purified from overnight cultures of ampicillin resistant an OD600 of 0.6 colonies using a Wizard Plus SV Miniprep DNA Purification System according to the manufacturer’s instructions (Promega).

ELISA. All samples were tested for antibodies to NPHV. Horse and donkey samples were also tested for antibodies to EpgV. High-bind 96-well ELISA plates (Greiner Bio-One) were coated overnight with BugBuster cell lysates (0.5 μg of recombinant protein in 100 μl of carbonate buffer) or an equivalent volume of protein isolated from cells infected with non-recombinant pGex-2t collected and processed in parallel with the NS3 and core proteins. Plates were washed with 250 μl of 1 % Tween 20/PBS (v/v), and coated wells were blocked with 150 μl of 3 % BSA/PBS (w/v) at room temperature for 2 h. After one round of washing, test samples were diluted 1:1000 in 100 μl of 5 % BSA/PBS (w/v), pre-incubated with 10 μg of recombinant pGex-2t cell lysate, and added to the wells and incubated for 1 h. The wells were washed six times with 250 μl of 1 % Tween 20/PBS (v/v) and incubated each time for 15 min, and then were incubated for 30 min with 100 μl of HRP-conjugated goat anti-horse immunoglobulin G antibody (Serotec) or the appropriate species specific anti-IgG diluted 1:4000 in 2 % BSA/PBS (w/v). After four rounds of washing, plates were developed by adding 70 μl of the HRP substrate (2,2’-azino-bis-3-ethylbenzothiazoline-6-sulfonic acid; Liquid Substrate System) to each well in accordance with the manufacturer’s protocol (Sigma). Plates were allowed to develop for 20–25 min and were read at 405 nm. The immunoreactivity of samples to core protein was compared and the mean control plus three times the sd was established as the cut-off. Samples confirmed as NPHV NS3 antibody positive were confirmed by second ELISA targeting NPHV core antigen (Tables S2, S3). EpgV antibody positive samples were confirmed by Western blot analyses.

Western blot analysis. Western blots were carried out to confirm all EpgV seropositive cases in the absence of a second confirmatory ELISA for EpgV. Whole-cell lysate and enriched protein samples were analysed on 10 %/16 % SDS-PAGE gels. The gels were either stained with Coomassie blue (Sigma) or blotted onto Protran BA 85 nitrocellulose membranes (Whatman) by use of a semidry electroblotter. Membranes and were blocked overnight in 4 % milk powder (MP)/0.05 % Tween 20/TBS (w/v). The membranes were then incubated at room temperature for 1 h with serum obtained from horses identified as EpgV seropositive and diluted 1:2000 in 4 % MP/0.05 % Tween 20/TBS (w/v). After six washes in 0.05 % Tween 20/TBS (v/v) for a total of 2 h, the membrane was incubated with HRP-conjugated goat anti-horse immunoglobulin G antibody (Serotec) diluted 1:35 000 in 4 % MP/0.05 % Tween 20/TBS (w/v). After four washes in 0.05 % Tween 20/TBS (v/v) for a total of 1 h, bound antibody was visualized using electrochemiluminescent prime Western blotting detection reagent in accordance with the manufacturer’s protocol (Amersham).
ACKNOWLEDGEMENTS

We thank clinicians at the Royal (Dick) School of Veterinary Studies, the University of Edinburgh, the Donkey Sanctuary, Devon, and Liphook Equine Hospital, Hampshire, for assistance in sourcing a wide variety of samples. This study was graciously supported by Google.org, the Skoll Foundation, and in particular the US Agency for International Development (USAID); USAID’s funding of this study was made possible by the generous support of the American people through the USAID Emerging Pandemic Threats PREDICT program. The content is the responsibility of the authors and does not necessarily reflect the views of USAID or the US Government. The funder had no role in the design of the study, data analysis, or preparation of the manuscript.

REFERENCES


