Development and application of viral whole genome sequencing for Severe Acute Respiratory Syndrome Coronavirus 2 and Hepatitis B virus

Sheila Frances Lumley

Trinity College

University of Oxford



A thesis submitted for the degree of

Doctor of Philosophy

Michaelmas Term 2024

Contents

Contents **2**
Abstract **6**
Acknowledgements **8**
List of Figures **11**
List of Tables **16**

Chapter 1: Introduction **20**
1.1. Context 20
1.2. Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) 21
1.2.1 SARS-CoV-2 Epidemiology 21
1.2.2 SARS-CoV-2 structure and genome 22
1.2.3 COVID-19 clinical presentation 26
1.2.4 SARS-CoV-2 antiviral strategies 26
1.2.5 SARS-CoV-2 transmission and containment measures 26
1.2.6 SARS-CoV-2 vaccination 27
1.2.7 Oxford University Hospitals healthcare worker cohort establishment 28
1.2.8 Studies on healthcare worker infection and immunity 29
1.3. Hepatitis B Virus (HBV) 30
1.3.1 HBV epidemiology 30
1.3.2 HBV global elimination targets 31
1.3.3 HBV clinical sequelae and disease progression 31
1.3.4 HBV structure and genome 32
1.3.5 HBV genetic diversity 34
1.3.6 HBV diagnosis and monitoring 36
1.3.7 HBV transmission 37
1.3.8 HBV prevention strategies 38
1.3.9 HBV treatment 39
1.3.10 Mechanisms of HBV drug resistance (HBVDR) 43
1.3.11 Definitions of HBVDR 46
1.3.12 HBVDR epidemiology 47
1.3.13 HBVDR testing 47
1.4. Viral whole genome sequencing as a translational tool 49
1.4.1 Applications of WGS to clinical and public health practice 49
1.4.2 Shared sequencing challenges for SARS-CoV-2 and HBV 49
1.4.3 Unique challenges with HBV WGS 50
1.4.4 Approaches to targeted viral sequencing 50
1.4.5 Approach to improving sensitivity of viral WGS methods 52
1.4.6 Sequencing platforms 52

1.5. Project rationale and thesis aims..... 54
1.5.1 SARS-CoV-2 projects and aims..... 54
1.5.2 HBV projects and aims..... 54

Chapter 2: Methods..... 56
2.1. SARS-CoV-2 methods..... 56
2.1.1 SARS-CoV-2 setting and samples..... 56
2.1.2 SARS-CoV-2 laboratory methods..... 58
2.1.3 SARS-CoV-2 bioinformatic pipelines and data analysis..... 59
2.2. Hepatitis B virus methods..... 60
2.2.1 HBV setting and samples..... 60
2.2.2 HBV laboratory methods..... 63
2.2.3 HBV bioinformatic pipelines and data analysis..... 75

Chapter 3: SARS-CoV-2 WGS development and application to infection prevention control..... 79
3.1. Availability of content..... 79
3.2. Introduction..... 80
3.3. Aims..... 80
3.4. Methods..... 81
3.4.1 Study design, setting and participants..... 81
3.4.2 SARS-CoV-2 WGS..... 82
3.4.3 Epidemiologic and genetic classification of community v.s. hospital acquisition 83
3.4.4 Outbreak analysis..... 85
3.5. Results..... 86
3.5.1 Overview of study population..... 86
3.5.2 Epidemiological classification..... 92
3.5.3 Can genomics help to confirm/exclude nosocomial acquisition?..... 97
3.5.4 Can genomics improve understanding of transmission patterns in hospitals?.. 99
3.6. Discussion..... 105

Chapter 4: Developing novel HBV whole genome sequencing methods..... 112
4.1. Availability of content..... 112
4.2. Background..... 112
4.3. Aims..... 114
4.4. HBV rolling circle amplification..... 115
4.4.1 Background..... 115
4.4.2 Methods and Results..... 117
4.4.3 Conclusion..... 126
4.5. Probe-based enrichment..... 128
4.5.1 Background..... 128

4.5.2 Methods..... 128
4.5.3 Results..... 129
4.5.4 Conclusion..... 136
4.6. HEP-TILE: tiled amplicon HBV WGS..... 137
4.6.1 Background..... 137
4.6.2 Methods..... 137
4.6.3 Results..... 138
4.6.4 Conclusion..... 144
4.7. Methods comparison HEP-TILE v.s. Probe based capture..... 145
4.7.1 Background..... 145
4.7.2 Methods..... 145
4.7.3 Results..... 146
4.7.4 Conclusion..... 147
4.8. Ultra low viral load samples..... 149
4.8.1 Background..... 149
4.8.2 Methods..... 149
4.8.3 Results..... 150
4.8.4 Conclusion..... 151
4.9. HBV WGS from dried blood spots..... 152
4.9.1 Background..... 152
4.9.2 Methods..... 152
4.9.3 Results..... 152
4.9.4 Conclusion..... 153
4.10. Discussion..... 154
4.10.1 Real world application of methods..... 154
4.10.2 Caveats and limitations..... 157
4.10.3 Challenges..... 158
4.10.4 Alternative methods..... 159
4.10.5 Conclusion..... 160

Chapter 5: HBV drug resistance..... 161
5.1. Availability of content..... 161
5.2. Background..... 161
5.3. Aims..... 164
5.4. Case report - factors influencing outcomes of NA therapy..... 164
5.5. A systematic review and meta-analysis of the risk of hepatitis B virus (HBV) resistance in people treated with entecavir or tenofovir..... 170
5.5.1 Background..... 170
5.5.2 Methods..... 170
5.5.3. Results..... 177
5.5.4 Discussion..... 192

5.6. WHO stakeholder analysis: knowledge gaps and plan for the field..... 194
5.6.1 Knowledge gaps.....194
5.6.2 Closing the research and data gaps in the HBVDR field..... 195
5.7. Conclusion and future work.....198

Chapter 6: Treatment eligibility and drug resistance in the WHO African region..... 199
6.1. Availability of content..... 199
6.2. Background.....199
6.3. Aims.....200
6.4. HBV treatment eligibility in the Uganda General Population Cohort.....202
6.4.1 Introduction..... 202
6.4.2 Methods..... 204
6.4.3 Results..... 205
6.4.4 Conclusion..... 213
6.5. Application of HBV WGS in four settings in the WHO African region..... 217
6.5.1 Introduction..... 217
6.5.2 Methods..... 218
6.5.3 Results..... 219
6.5.4 Conclusion..... 239
6.6. Discussion.....241

Chapter 7: Discussion.....243
7.1 Introduction..... 243
7.2 Summary of work, impact and limitations.....243
7.2.1 SARS-CoV-2: WGS insights into nosocomial transmission..... 243
7.2.2 Improved techniques for HBV WGS.....245
7.2.3 HBV drug resistance knowledge gaps and systematic review.....246
7.2.4 Application of new WHO HBV treatment guidelines in Uganda.....247
7.2.5 HBV WGS in the WHO African region..... 248
7.3 Challenges with SARS-CoV-2 and HBV research..... 249
7.4 Future recommendations and next steps.....250

Appendices.....252
Appendix 1: Publications and presentations.....252
Appendix 2: Sample metadata.....254
Appendix 3: Supplementary material for Chapter 5.....266

References.....283

Abstract

During my DPhil, I set out to develop and apply sequencing techniques to Hepatitis B Virus (HBV), a pathogen which causes over a million deaths a year and for which molecular epidemiology is poorly resolved. However, this work was eclipsed by the COVID-19 pandemic caused by Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) which rapidly became the world's biggest viral killer. I rapidly applied the tools and approaches planned for endemic HBV to tackle pandemic SARS-CoV, implementing whole genome sequencing (WGS) techniques as a translational tool.

I established SARS-CoV-2 WGS in my local hospital and explored how sequencing can help track the spread of SARS-CoV-2 in healthcare settings. By combining WGS with epidemiological data, I showed that most hospital-acquired infections result from superspreading events, providing insights to inform infection prevention control (IPC) practice.

Returning to work on HBV, I optimised WGS tools, developing methods for HBV WGS, including the easy to use, open-access HEP-TILE protocol. To explore the clinical need for sequencing, I conducted a systematic review and meta-analysis to determine the epidemiology of drug resistance, demonstrating the high genetic barrier for tenofovir resistance compared to entecavir. Using HEP-TILE, I generated HBV genomes from adults with CHB in Uganda, Kenya, South Africa and the Democratic Republic of Congo. I described new HBV strains including recombinants and established the prevalence of resistance associated mutations (RAMs). I also discovered a potential link between a specific HBV polymorphism (S13T) and liver disease (fibrosis/cirrhosis). To further develop insights

into HBV treatment, I explored the application of HBV treatment guidelines in Uganda, highlighting the limitations of current non-invasive tests.

My thesis demonstrates how viral WGS can be applied to different viruses to enhance public health management. Globally, sequencing expertise and infrastructure gained during the COVID-19 pandemic can be applied to HBV, to support progress towards 2030 elimination targets for viral hepatitis.

Acknowledgements

I would firstly like to acknowledge my fantastic supervisors Professor Philippa Matthews and Professor Ellie Barnes for their guidance and expertise. I feel incredibly lucky to have had Philippa as my primary supervisor and I will always be grateful to her for her unwavering encouragement through my academic clinical fellowship and DPhil. It has been challenging having to pivot our work from HBV to SARS-CoV-2 and back again, but she has supported me to follow my interests, provided incredible opportunities for international collaboration and led with fairness, kindness and compassion. Ellie has provided perspective, mentorship and continues to encourage and support my ambition to be a clinical academic. Both have contributed greatly to my progress throughout my DPhil and continue to inspire me.

I would like to thank all of the Matthews group members past and present, for their academic discussions, expertise and friendship, it has been a real privilege to work together. Particular thanks to Anna McNaughton for her ongoing help and collaboration. During my SARS-CoV-2 work I had the opportunity to work in the Modernising Medical Microbiology group and NIHR Health Protection Research Unit in Healthcare Associated Infections led by Derrick Crook and Tim Peto at the John Radcliffe Hospital, Oxford. I am very appreciative of the support and freedom they afforded me in establishing SARS-CoV-2 sequencing and to the many group members who I worked with. Special thanks also go to David Eyre, Katie Jeffery, Tim Walker and Nicole Stoesser, the biomedical scientists and infection prevention control team at Oxford University Hospitals, whom I worked with closely during my SARS-CoV-2 work. Also thanks to Denise O'Donnell and the large team of 100+ medical students and allied healthcare professionals who volunteered to work at the asymptomatic healthcare worker testing clinics.

I would like to acknowledge my collaborators; Gavin Screaton's lab for SARS-CoV-2 collaborations, Azim Ansari's group and Paolo Piazza's team for HBV collaborations. I would like to thank Josh Quick, Nick Loman and their teams at the University of Birmingham, with special thanks to Chris Kent, bioinformatician and PhD student in the group. I would also like to thank the wider Medawar community who have supported me throughout, Paul Klenerman for his mentorship, the Goulder group for lending me some much needed bench space, Cathy for keeping everyone in check, my office-mates from the Barnes group who have provided camaraderie and entertainment, and all the friends I've made over the last 10 years who have made the experience enjoyable.

I have really enjoyed working with collaborators and team members in South Africa, Kenya and Uganda. Thanks to the huge team who coordinates the Uganda Liver Disease Study, it is an incredible set up and I am very grateful I had the chance to spend time working with you there. Special thanks to Robert Newton, Bernard Mpairwe, Beatrice Kimono, Ronald Maponga, Kevin Ojambo and Moses Kwizera in addition to the research teams at UVRI who have been so helpful and accommodating. Thanks to Louise Downs and Dorcas Okanda in Kenya, I am so grateful that I could come and see your incredible work first hand.

This work would not have been possible without the individuals who selflessly took part in these studies. I am also thankful to the Wellcome Trust for providing me with a Clinical Research Training Fellowship grant to fund this research.

Thank you to everyone at Green Templeton College, from the porters and pantry staff to my colleagues and students, for being such a wonderful community. Thank you to my amazing friends for riding the highs and lows of academia, parenthood and pandemic life with me, and keeping everything in perspective. Finally, thank you to my family, my husband Michael,

children Oliver and Clara and my parents Sarah and Philip, for their love, support, fun and adventures; without you this work would not have been possible.

List of Figures

Figure 1.1. WHO COVID dashboard showing total and weekly reported COVID-19 deaths worldwide

Figure 1.2. Schematic diagrams of the SARS-CoV-2 virus particle and genome

Figure 1.3. Genomic epidemiology of SARS-CoV-2 from 2020 - 2024

Figure 1.4. Prevalent cases of CHB by WHO region 2022

Figure 1.5. Schematic representation of HBV particles

Figure 1.6. Schematic representation of the structure of HBV genomic DNA, RNAs, and proteins

Figure 1.7. Genomic-length maximum-likelihood phylogenetic tree of HBV genotype, subgenotype and clade reference strains

Figure 1.8. Schematic to illustrate genetic basis of HBVDR

Figure 1.9. Comparison of current HBV resistance genotyping interpretation algorithms

Figure 1.10. Bimodal distribution of HBV VL

Figure 2.1. Example HBV plasmid construct with a 1.3mer HBV sequence inserted into the vector pUC57

Figure 2.2. Schematic to show the process of probe design for HBV probe-based capture

Figure 2.3. Schematic to show the design of multiplex tiled PCR for the circular HBV genome

Figure 2.4. Modifications to HEP-TILE primer positions to take into account HBV genome structure

Figure 3.1. Epidemic curves showing daily count of UK confirmed SARS-CoV-2 cases by specimen date

Figure 3.2. Epidemic curves showing number of new SARS-CoV-2 PCR staff and inpatient positives at Oxford University Hospitals from 17 November 2020 - 5 January 2021 based on clinical microbiology laboratory diagnosis by PCR

Figure 3.3. Flow diagram showing numbers of samples available for sequencing and being successfully sequenced

Figure 3.4. SARS-CoV-2 outbreaks at Oxford University Hospitals between 17 November 2020 and 5 January 2021 containing at least one definite or probable nosocomial case.

Figure 3.5. Timing and size of SARS-CoV-2 genetic clusters at Oxford University Hospitals between 17 November 2020 and 5 January 2021

Figure 3.6. Example phylogeny demonstrating a superspreading event and recurrent introductions on a single ward at hospital “B” (Oxford University Hospitals)

Figure 4.1. Overview of three approaches to targeted viral enrichment for HBV whole genome sequencing

Figure 4.2. Flow diagram to show how rolling circle amplification, library preparation and bioinformatic processing can be used to generate error-corrected full length HBV genomes

Figure 4.3. Impact of modifications to bead clean up on DNA size selection

Figure 4.4. Impact of host depletion on HBV sequencing outputs in samples enriched by rolling circle amplification

Figure 4.5. Optimisation of rolling circle amplification protocol enzymes and primers

Figure 4.6. Strategies for generating a fully circular HBV template for rolling circle amplification

Figure 4.7. Flow chart summarising sample sets for enrichment with probe based capture

Figure 4.8. Comparison of number of HBV reads per million sequenced with/without probe-based capture

Figure 4.9. Sequencing outputs for HBV WGS using capture for viral enrichment

Figure 4.10. HBV genome coverage for mock samples with HBV genotypes (A-J) enriched using HBV probe based capture

Figure 4.11. Sample sets for tiled amplicon (HEP-TILE) enrichment optimisation

Figure 4.12. Sequencing outputs for HBV WGS using amplicon based target enrichment (HEP-TILE)

Figure 4.13. Relationship between HBV genotype and performance of HEP-TILE Nanopore workflow

Figure 4.14. HBV genome coverage for mock samples of HBV genotypes (A-J) enriched using HEP-TILE and sequenced on (A) nanopore and (B) illumina platforms

Figure 4.15. Overview of sample sets sequenced by 3 methods in sections 5 and 6, used in analysis to compare outputs of different platforms

Figure 4.16. Sequencing metrics from low viral load samples extracted using high volume extraction

Figure 4.17. HBV genome coverage using a tiled amplicon protocol from the HOVER-HBV study in Democratic Republic of Congo

Figure 5.1. Summary of treatment and laboratory results from a case report of an individual with non-suppression of HBV VL despite being prescribed antiviral therapy

Figure 5.2. Gold standard data flow to determine risk of clinical and genotypic resistance.

Figure 5.3. PRISMA flowchart for HBVDR systematic review

Figure 5.4. Characteristics of 62 studies included in our systematic review and meta analysis to determine the risk of TFV and ETV resistance in individuals receiving treatment for CHB infection

Figure 5.5. Meta-analysis of risk of ETV resistance over time in NA naive individuals

Figure 5.6. Meta-analysis of risk of ETV resistance over time in NA experienced individuals

Figure 5.7. Meta-analysis of risk of TFV resistance over time in NA naive individuals

Figure 5.8. Meta-analysis of risk of TFV resistance over time in NA experienced individuals

Figure 6.1. HBV treatment eligibility criteria from the WHO guidelines for the prevention, diagnosis, care and treatment for people with chronic hepatitis B infection, March 2024

Figure 6.2. Map of Africa showing locations of study settings in Kenya, South Africa, Democratic Republic of the Congo and Uganda

Figure 6.3. Flow diagram to show the timeline of recruitment, testing and follow up of Uganda liver disease study (ULiDS) HBV cohort between 2011 - 2023 in Kalungu District, Uganda

Figure 6.4. Eligibility for HBV treatment in Uganda liver disease study (ULiDS) HBV cohort based on WHO 2024 treatment guidelines

Figure 6.5. Predictive value of routinely available biomarkers for detecting fibrosis or cirrhosis in treatment-naive individuals with CHB mono-infection in the Uganda liver disease study (ULiDS) 2023, using TE as the gold standard

Figure 6.6. HBV VL in 2011 and 2023 in HBV mono-infected untreated individuals from the Uganda liver disease study (ULiDS)

Figure 6.7. Application of WGS to four settings in WHO African region leads to shared insights into genotypes, recombinants, known and novel resistance associated mutations (RAMs) and transmission

Figure 6.8. Percent HBV genome coverage split by HBV VL in four settings in the WHO African region

Figure 6.9. Maximum likelihood phylogenetic tree of HBV WGS from Uganda with genotype A, D and E reference strains

Figure 6.10. Maximum likelihood phylogenetic tree of HBV WGS from South Africa with genotype A, C, D and E reference strains

Figure 6.11. Presence of established and putative RAMs in 110 treatment naive and 6 treatment experienced individuals with CHB

Figure 6.12. Polymorphic sites in HBV RT sequences from six individuals with virological breakthrough on tenofovir

Figure 6.13. Maximum likelihood phylogenetic tree showing genotype and household clustering in the Democratic Republic of Congo with genotype A, D and E reference strains

Figure 6.14. Maximum likelihood phylogenetic trees showing genotype and household clustering in Kenya with genotype A, D and E reference strains

Figure 6.15. Family trees for two households in Kenya highlighted in the phylogenetic tree in Figure 6.14

Supplementary figure 1. Summary of risk of bias scores for included studies

Supplementary figure 2. Relationship between clinical and methodological sources of heterogeneity

Supplementary figure 3. Funnel plots to investigate publication bias

List of Tables

Table 1.1. Nucleos(t)ide analogue agents active against HBV

Table 1.2. Comparison of international HBV guidelines for defining, monitoring and managing treatment failure

Table 1.3. Existing and proposed definitions pertaining to HBV drug resistance (HBVDR)

Table 1.4. Comparison of viral enrichment methods

Table 1.5. Comparison of strengths and limitations of Illumina and Nanopore sequencing platforms

Table 2.1. Description and ethical approvals of HBV cohorts in the United Kingdom and Africa

Table 2.2. Assays used for HBV VL quantification across settings

Table 2.3. Methods for nucleic acid extraction from samples obtained from individuals with CHB across settings

Table 2.4. Rolling circle amplification primers from Margeridon et al.

Table 2.5. Full length gPCR primers from Gunther et al.

Table 2.6. Methods for HBV enrichment and library preparation

Table 2.7. RAMs which individually or in combination have been associated with HBV drug resistance

Table 3.1. Characteristics of 803 inpatients testing SARS-CoV-2 PCR positive between 17 November 2020 and 5 January 2021 at Oxford University Hospitals

Table 3.2. SARS-CoV-2 sequencing pass rates by the standard epidemiological nosocomial classification and Thermofisher Ct value

Table 3.3. Demographics of inpatients testing PCR positive for SARS-CoV-2 infection at Oxford University Hospitals between 17 November 2020 and 5 January 2021, representing those for whom SARS-CoV-2 was sequenced vs non-sequenced

Table 3.4. Epidemiological and genomic evidence for nosocomial acquisition of SARS-CoV-2 based on classification of 803 PCR-positive hospital in-patients

Table 4.1. Viral load and sequencing output for 20 plasma samples from adults with CHB enriched using 'Takara-capture' protocol and sequenced by Illumina

Table 4.2. Sequencing outputs for clinical HBV samples sequenced by Takara-capture

Table 4.3. Sequencing metrics for mock HBV samples sequenced with/without probe based capture and sequenced on Illumina

Table 4.4. Sequencing outputs for clinical HBV samples sequenced by HEP-TILE

Table 4.5. Sequencing metrics for mock HBV samples for HEP-TILE Nanopore and HEP-TILE Illumina workflows

Table 4.6. Comparison of sequencing errors when sequencing mock samples using three workflows

Table 4.7. Summary of four HBV WGS pipelines, comparing time, cost and outputs

Table 5.1. Factors that potentially underlie inadequate virological response to therapy for hepatitis B virus (HBV) with nucleos(t)ide analogue (NA) agents

Table 5.2. HBV drug resistance report from a clinical diagnostic virology laboratory, based on a serum sample submitted during VL rebound (year 25)

Table 5.3. Resistance associated mutations (RAMs) listed in EASL guidance, Genotype A/G consensus and polymorphisms identified in HBV isolated from this individual

Table 5.4. Suggestions for closing research gaps in the HBVDR field

Table 6.1. WHO HBV treatment criteria 2024 mapped against data available in Uganda Liver Disease Study

Table 6.2. Uganda Liver Disease Study (ULiDS) cohort characteristics at baseline (2011) and in those followed up (2023)

Table 6.3. Details of the Hepatitis B virus whole genome sequencing protocol used for different cohorts

Table 6.4. Comparison of HBV WGS coverage across four settings in the WHO African region

Table 6.5. Distribution of HBV genotypes across four settings in the WHO African region

Table 6.6. Amino acid residues found at established and putative RAM sites in HBV reference strains

Table 6.7. Prevalence of combinations of RAMs associated with reduced or intermediate susceptibility to NAs in four settings

Table 6.8. RT polymorphisms in six individuals with VBT on tenofovir

Table 6.9. Frequency of HBV polymorphisms identified in consensus sequences derived by Nanopore sequencing of NA-experienced and NA-naive individuals with genotype A infection

Table 6.10. Frequency of HBV polymorphisms previously associated with cirrhosis identified in consensus sequences derived by Nanopore sequencing among patients with and without a Transient Elastography score > 7kPa (F2 fibrosis)

Supplementary Table 1. HBV samples used for rolling circle amplification

Supplementary Table 2. HBV samples used for takara library prep and capture

Supplementary Table 3. HBV samples used for NEB library prep and capture

Supplementary Table 4:. HBV samples used for HEP-TILE proof of principle work

Supplementary Table 5. HBV samples used for ultra low VL work

Supplementary Table 6. HBV samples from Democratic Republic of the Congo

Supplementary Table 7. HBV samples from Uganda

Supplementary Table 8. HBV samples from Kenya

Supplementary Table 9. PRISMA checklist for systematic review and meta-analysis of tenofovir and entecavir resistance

Supplementary Table 10. Analysis of resistance criteria for TFV papers included in HBVDR systematic review and meta-analysis

Supplementary Table 11. Characteristics of 62 studies included in the systematic review

Supplementary Table 12. Sensitivity analysis for Experienced/Tenofovir

Supplementary Table 13. Sensitivity analysis for Naive/Entecavir

Supplementary Table 14. Sensitivity analysis for Experienced/Entecavir

Chapter 1: Introduction

1.1. Context

Viral pathogens have a profound impact on global health, contributing significantly to both morbidity and mortality. Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) and Hepatitis B Virus (HBV) are two such examples. Since HBV was first identified in the 1940s[1] it has presented a persistent global health challenge; in contrast, SARS-CoV-2 emerged in December 2019[2], causing a global pandemic.

I started my Wellcome DPhil fellowship in February 2020, with an initial focus on HBV, however after six weeks, the laboratories at Oxford University were closed due to the SARS-CoV-2 pandemic and I was re-deployed to full time National Health Service clinical work at Oxford University Hospitals (OUH). In this context, I became involved in translational SARS-CoV-2 research for 18 months and returned to HBV research in January 2023 following a year of parental leave. My thesis therefore contains a combination of SARS-CoV-2 and HBV research. To maintain a cohesive thesis within the word limit, I provide a summary of SARS-CoV-2 outputs in **Appendix 1** and dedicate one results chapter (**Chapter 3**) to the SARS-CoV-2 sequencing outputs from the OUH healthcare worker cohort which closely align with my subsequent work on HBV.

In my thesis, I consider what data we need to inform public health management of SARS-CoV-2 and HBV and address some key knowledge gaps. I develop and deploy whole genome sequencing (WGS) methods for both pathogens. In the case of SARS-CoV-2, I explore the utility of WGS as a translational clinical tool to inform infection prevention control (IPC) in the hospital setting (**Chapter 3**). In the case of HBV, I consider WGS in the context of

HBV drug resistance (HBVDR). I develop WGS methods to monitor HBVDR (**Chapter 4**), I investigate what is required in the context of World Health Organisation (WHO) expanded treatment guidelines to mitigate against potential HBVDR and fill a key evidence gap with a meta-analysis of HBVDR (**Chapter 5**). I then consider the impact and issues surrounding the application of new WHO treatment guidelines in a rural setting in Uganda, and finish by rolling out HBV WGS across four settings in Africa and analysing preliminary HBVDR data (**Chapter 6**).

This introductory chapter gives a broad overview of key concepts relevant to my thesis. More detailed chapter-specific background is included at the start of respective chapters.

1.2. Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2)

1.2.1 SARS-CoV-2 Epidemiology

SARS-CoV-2 was first identified in China in December 2019 when several health facilities in Wuhan reported clusters of patients with pneumonia of unknown cause[3]. Most initial cases were linked to a market in Wuhan selling seafood and live animals. Within a few months SARS-CoV-2 had spread globally. A global health emergency was declared by the WHO on 30th January 2020, at the same time that the first cases were diagnosed in the UK[4]. A pandemic was declared by the WHO in March 2020, with millions of cases reported globally by the end of 2020. By the end of 2024 SARS-CoV-2 had caused over 7 million deaths[5] (**Figure 1.1**).

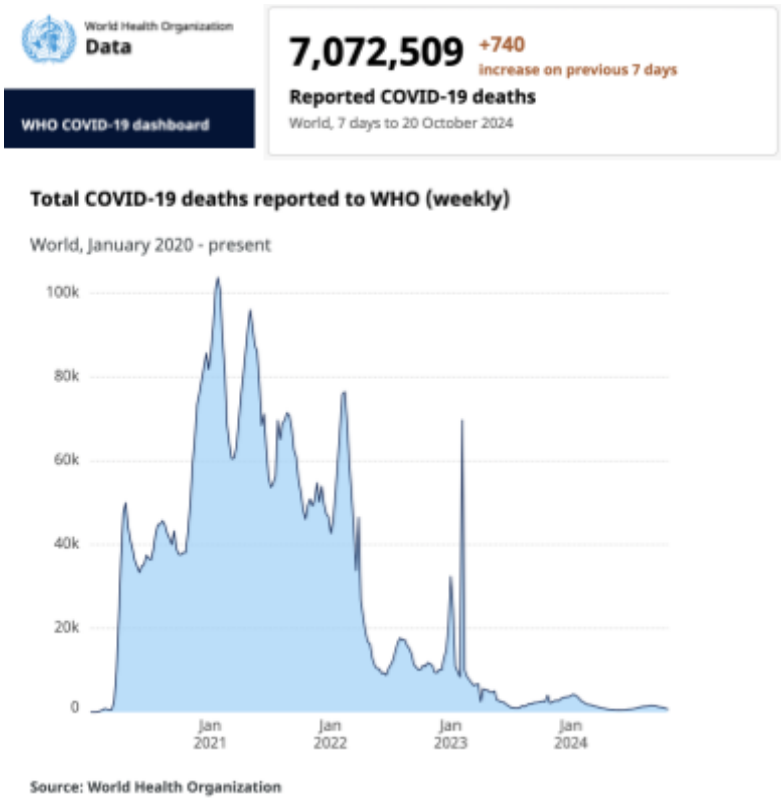


Figure 1.1. WHO COVID dashboard showing the total and weekly reported deaths due to COVID-19 worldwide. WHO dashboard accessed 7th November 2024[6]

1.2.2 SARS-CoV-2 structure and genome

SARS-CoV-2 is part of the Coronaviridae family, classified as a betacoronavirus. It is most closely related to a bat coronavirus (bats are natural hosts of betacoronaviruses) and shares 79% genome sequence identity with SARS-CoV and 50% with MERS-CoV. It has a single-stranded, positive-sense RNA genome of approximately 29.9 kilobases in length (**Figure 1.2**). The genome encodes several structural proteins, including the spike, envelope, membrane, and nucleocapsid proteins, as well as a number of non-structural proteins involved in viral replication and transcription (**Figure 1.2**).

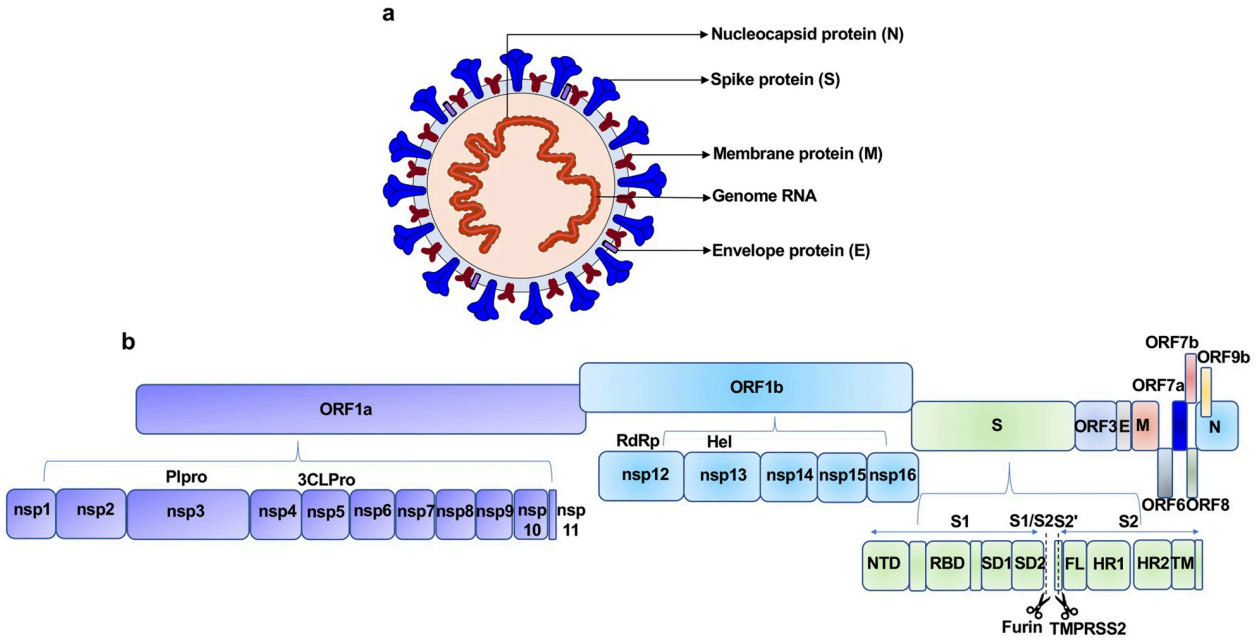


Figure 1.2. Schematic diagrams of the SARS-CoV-2 virus particle and genome.

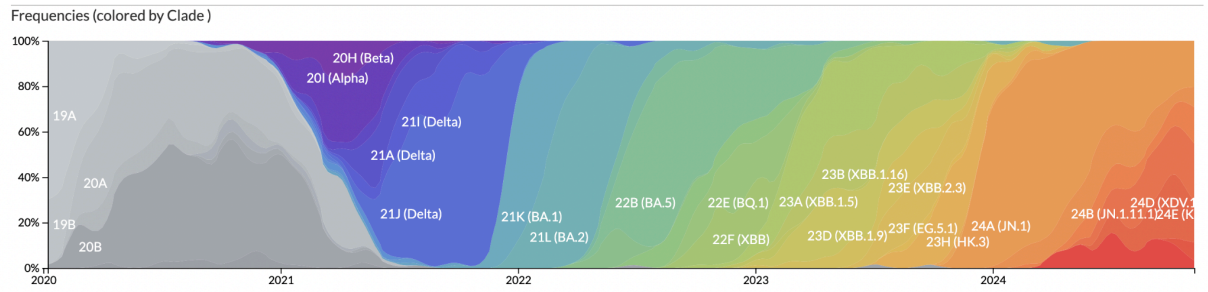
a) Four structural proteins of SARS-CoV-2 include Spike protein (S), Membrane protein (M), Nucleocapsid protein (N), and Envelope protein (E).

b) The genome includes ORF1a-ORF1b-S-ORF3-E-M-ORF6-ORF7 (7a and 7b) -ORF8-ORF9b-N in order. Sixteen nonstructural proteins (nsp1–11, 12–16) are encoded by ORF1a and ORF1b, respectively, and six accessory proteins were delineated. Plpro papain like protease, 3CLPro 3C-like proteinase, RdRp RNA-dependent RNA polymerase, Hel Helicase, S encodes NTD N-terminal domain, RBD receptor-binding domain, SD1 subdomain 1, SD2 subdomain 2, FL fusion loop, HR1 heptad repeat 1, HR2 heptad repeat 2, TM transmembrane domain. Dotted line indicates S1/S2 and S2′ site cleavage by Furin and TMPRSS2. Reproduced from[7] under a CC.BY license.

WGS of SARS-CoV-2 has occurred at an unprecedented scale; more than 950,000 full genome sequences were available in the Global Initiative on Sharing All Influenza Data (GISAID) database by April 2021[8]. WGS has allowed us to inform vaccine design, monitor transmission, and identify and characterise variants of concern (VOCs), such as Alpha (first identified in December 2020), Delta, and Omicron (**Figure 1.3**). SARS-COV-2 lineages are

assigned using a dynamic nomenclature based on a maximum likelihood phylogenetic tree, where each lineage must show phylogenetic evidence of emergence in a new geographic population with ongoing transmission[9].

A



B

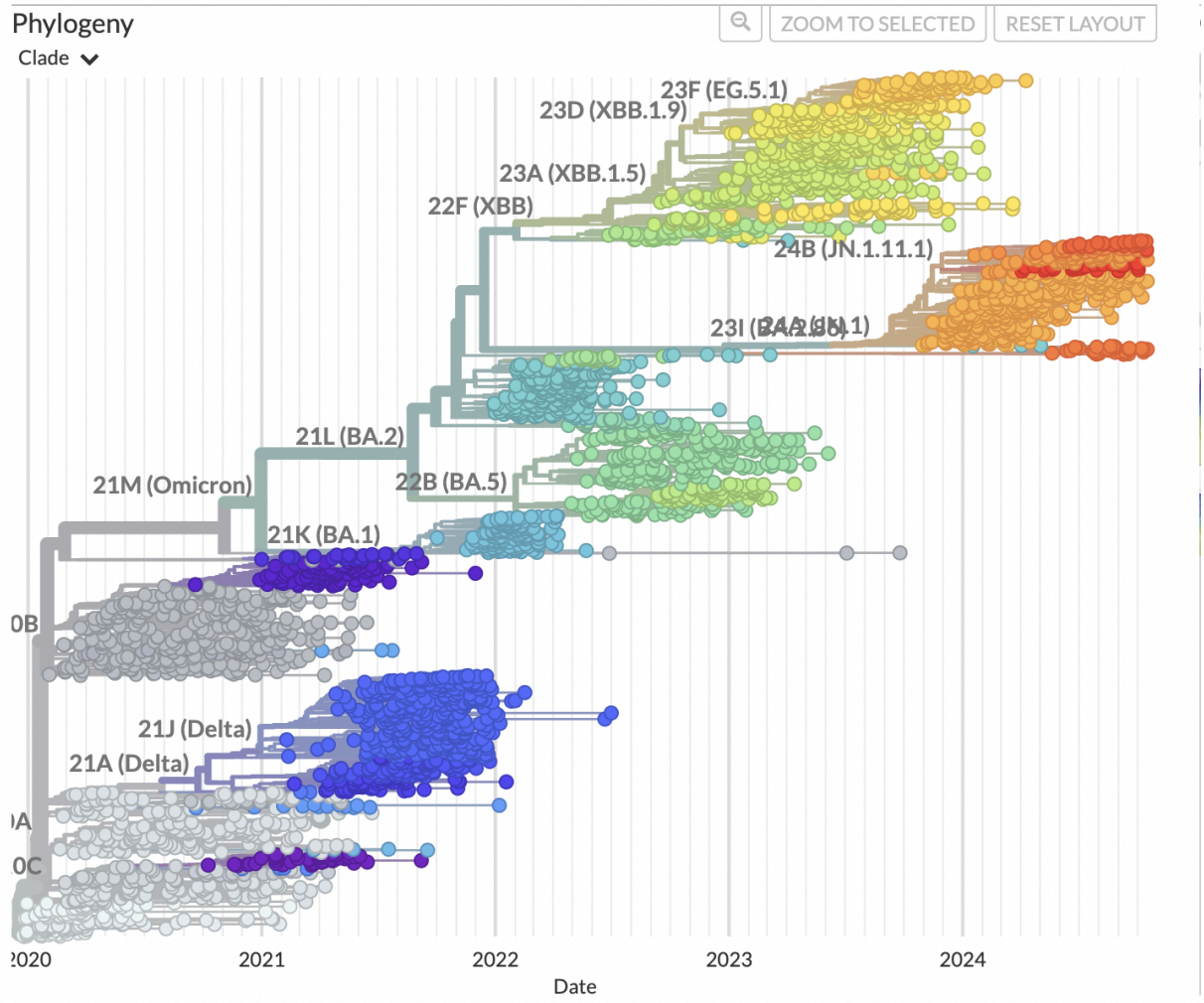


Figure 1.3. Genomic epidemiology of SARS-CoV-2 from 2020 - 2024.
A) Frequency of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) lineages over time. B) Phylogeny showing evolutionary relationships of circulating SARS-CoV-2 viruses. Grey = Pre-VOC. Reproduced from Nextstrain.org under a CC.BY licence[10]

1.2.3 COVID-19 clinical presentation

Coronavirus disease 2019 (COVID-19), the disease caused by SARS-CoV-2, presents with a wide range of symptoms, from asymptomatic or mild cases to severe illness requiring hospitalization[11]. Common symptoms include fever, cough, dyspnoea, anosmia, ageusia and fatigue, although many patients also experience non-respiratory symptoms. Severe cases can progress to acute respiratory distress syndrome (ARDS), multi-organ failure, and death. Long COVID, characterised by the persistence of symptoms for more than 3 months after initial infection, has led to a significant burden of ongoing population morbidity[12].

1.2.4 SARS-CoV-2 antiviral strategies

Dexamethasone was the first treatment approved for use in severe COVID-19 in June 2020. Antivirals were subsequently developed, molnupiravir (Lagevrio) a ribonucleoside analogue, was the first to be approved by MHRA in November 2021[13]. The UK National Institute for Clinical Excellence (NICE) currently recommends nirmatrelvir plus ritonavir (Paxlovid) a protease inhibitor, remdesivir (Veklury) an RNA polymerase inhibitor, and sotrovimab (Xevudy) a neutralising monoclonal antibody[14] as first line agents with choice dependant on disease severity.

1.2.5 SARS-CoV-2 transmission and containment measures

SARS-CoV-2 transmission occurs by several routes, including respiratory droplets, aerosols, and contact with contaminated surfaces. The incubation period of SARS-CoV-2 typically ranges from 2 to 14 days[15]. The ability of infected individuals to transmit the virus during the pre-symptomatic phase and during mild/asymptomatic infection, combined with the wide range of incubation periods, has made containment challenging. Containment methods used to control the virus spread ranged from nationwide lockdowns, international travel bans and

school closures, to curfews, social distancing measures, restrictions on gatherings, mandatory mask wearing, testing, contact tracing and isolation.

At the start of the COVID-19 pandemic, when the population lacked immunity from natural infection or vaccination, morbidity and mortality rates were high, especially in healthcare settings where nosocomial SARS-CoV-2 infection had crude patient mortality rates between 25% and 35%[16,17]. Extensive IPC measures were implemented in healthcare settings including use of personal protective equipment (PPE), hand hygiene, staff and patient testing, environmental cleaning, improved ventilation, patient isolation and visitor restrictions[18]. In the early stages of the pandemic, shortages of PPE and lack of adequate testing capacity led to high rates of infection among healthcare workers and inpatients[19]. However, as the pandemic progressed, global efforts to improve supply chains and IPC protocols, and the development of population immunity acquired by natural infection and vaccination, helped reduce these risks.

1.2.6 SARS-CoV-2 vaccination

The rapid development of vaccines against SARS-CoV-2 has been crucial in the global effort to control the COVID-19 pandemic. The work presented in my thesis spans the start of the UK vaccination campaign in December 2020, i.e. before widespread vaccine-based immunity. The first SARS-CoV-2 vaccines were authorized for emergency use in December 2020, the Pfizer/BioNTech (mRNA) and AstraZeneca/Oxford vaccines (adenovirus vector, developed by the Jenner Institute and Oxford Vaccine Group at the University of Oxford) were the most widely used to protect the UK population in 2020/21. These vaccines demonstrated high efficacy (70-90%) in preventing symptomatic COVID-19 and severe outcomes in clinical trials[20,21].

1.2.7 Oxford University Hospitals healthcare worker cohort establishment

Despite a national lockdown in the UK in March 2020 to mitigate the impact of the rapidly evolving COVID-19 pandemic, COVID-19 incidence amongst healthcare workers (HCWs) was recognised to be higher than in the general population[22,23]. In April 2020, evidence was conflicting about the greatest risks posed to HCWs, and knowledge about the duration and dynamics of immune responses after natural infection were limited. Longitudinal cohorts with baseline serology were required to determine the assay- and time-dependent dynamics of post-infection and post-vaccine antibody responses, how these differ between individuals e.g. by age, gender, ethnicity, and how these findings relate to protection from (re)-infection. Therefore I was involved in establishing a longitudinal cohort of OUH staff, to track infection and immunity, generating SARS-CoV-2 sequences and antibody data to advance our understanding of the clinical and molecular epidemiology of SARS-CoV-2 and inform surveillance, diagnostics, and IPC.

In April 2020, I led the team providing voluntary asymptomatic staff screening at OUH. I coordinated the rapid scale up of testing clinics, through which we provided PCR and antibody testing for over 9000 HCWs in the first month. I was involved with logistics, procurement, clinic staff recruitment, staff training and communications. I initially established clinics at the John Radcliffe hospital site, then also set up and ran a satellite testing unit at the Churchill hospital (both close to central Oxford). With over 14,000 staff enrolled, and many engaged in active follow-up, this large longitudinal cohort put us in an excellent position to address questions pertaining to transmission and immunity. These data were shared with the national SIREN HCW project, a prospective cohort study of staff working in publicly funded hospitals in the UK[24,25] .

1.2.8 Studies on healthcare worker infection and immunity

I published a series of first- and middle-author papers from the OUH HCW cohort, I provide a summary of the outputs in **Appendix 1**, and dedicate one results chapter (**Chapter 3**) to the SARS-CoV-2 sequencing outputs from the OUH HCW cohort.

1.3. Hepatitis B Virus (HBV)

1.3.1 HBV epidemiology

Chronic HBV infection (CHB) is a global public health challenge, affecting an estimated 254 million people worldwide, despite the availability of effective vaccines[26]. CHB continues to be a significant cause of liver-related morbidity and mortality, accounting for approximately 1.1 million deaths in 2022[27]. The WHO African region has among the highest incidence and mortality rates from liver disease in the world due to the high prevalence of viral hepatitis, alcohol and non-alcoholic fatty liver disease[28]. HBV infection is a significant public health challenge in this region, with an estimated 65 million people chronically infected and approximately 270,000 HBV-related deaths occurring in Africa each year (**Figure 1.4**). The WHO 'Action for Access in Low and Middle Income Countries (LMIC)' report published in 2024 highlights that 66% of new infections occur in the WHO African region[29].

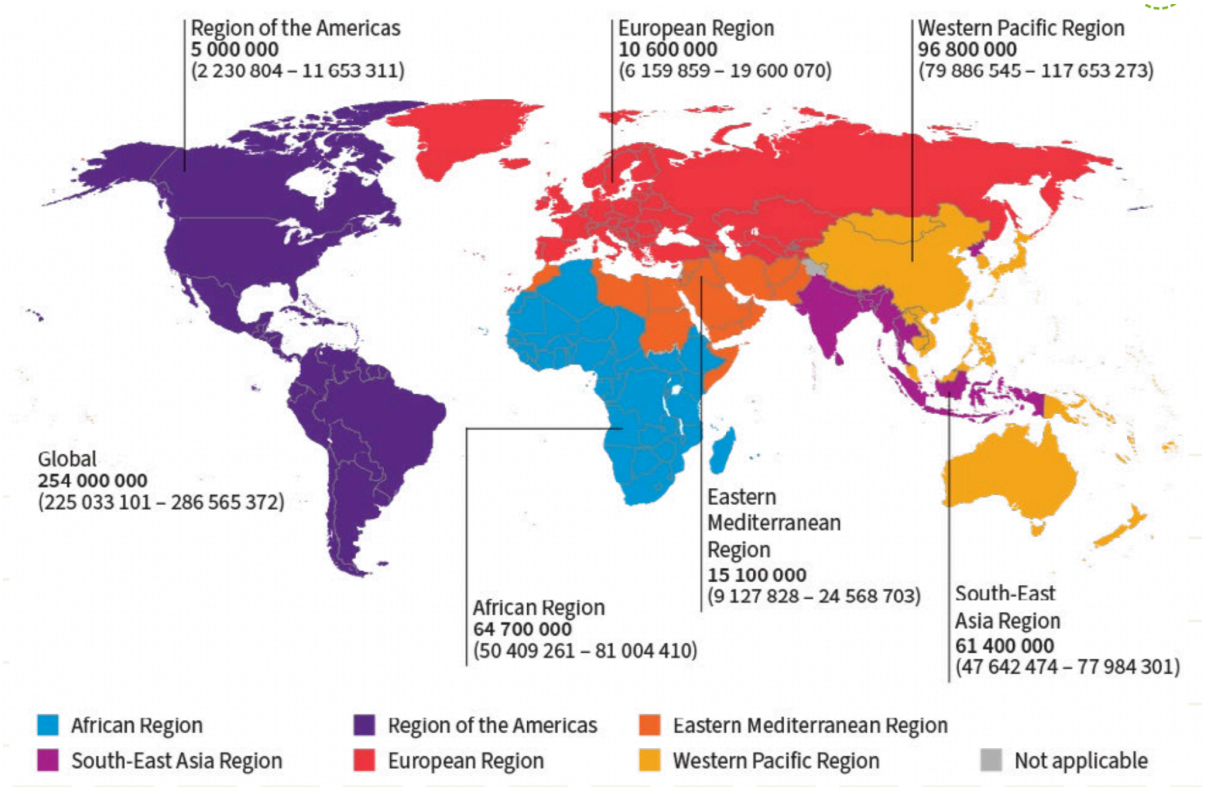


Figure 1.4. Prevalent cases of CHB by WHO region 2022. Figure reproduced under CC BY licence[30] .

1.3.2 HBV global elimination targets

The United Nations has recognized HBV elimination as a priority, including it in the Sustainable Development Goals (Target 3.3). In alignment with these goals, the WHO published the Global Health Sector Strategy on viral hepatitis in 2016[31], setting out to 'eliminate HBV as a global public health threat by the year 2030', with specific targets including aiming for a 90% reduction in HBV incidence and a 65% reduction in mortality by 2030. To achieve these objectives, the strategy outlines service coverage targets, including reaching 90% childhood vaccination coverage, 90% testing coverage and 80% treatment coverage by 2030. However, progress toward these targets has been slow, particularly in LMICs where limited healthcare infrastructure hampers access to essential testing and treatment services, where resourcing, awareness and advocacy have been chronically neglected, and where populations are inadequately represented by research and clinical trials[32,33].

1.3.3 HBV clinical sequelae and disease progression

The risk of progression of HBV infection to CHB is inversely related to the age at which infection occurs. Perinatal and early childhood infections are more likely to lead to chronic disease (90-95% risk of chronicity), whereas most healthy adults clear acute infection spontaneously (90-95% chance of clearance). CHB is a major cause of liver-related morbidity and mortality. Without treatment, up to 25% of individuals with chronic HBV infection will develop cirrhosis, liver failure, or hepatocellular carcinoma (HCC) over their lifetime[34]. In Africa, HCC is now the leading cause of cancer deaths in males under age 50[35]. There is also increasing recognition of a burden of other health consequences such as fatigue, anxiety and depression, highlighting the profound psychosocial impacts of living with a highly stigmatised infection[36,37].

1.3.4 HBV structure and genome

HBV is a small, enveloped DNA virus in the Hepadnaviridae family. The infectious virion (Dane particle) consists of a lipid membrane with three HBV surface antigens (HBsAg), large (L-HBs), middle (M-HBs), and small (S-HBs), that surround a nucleocapsid composed of hepatitis B core protein (HBc), viral polymerase (Pol), and viral DNA. A range of other non-infectious HBV particles of varying shape and size are also found in the serum of infected individuals (**Figure 1.5**).

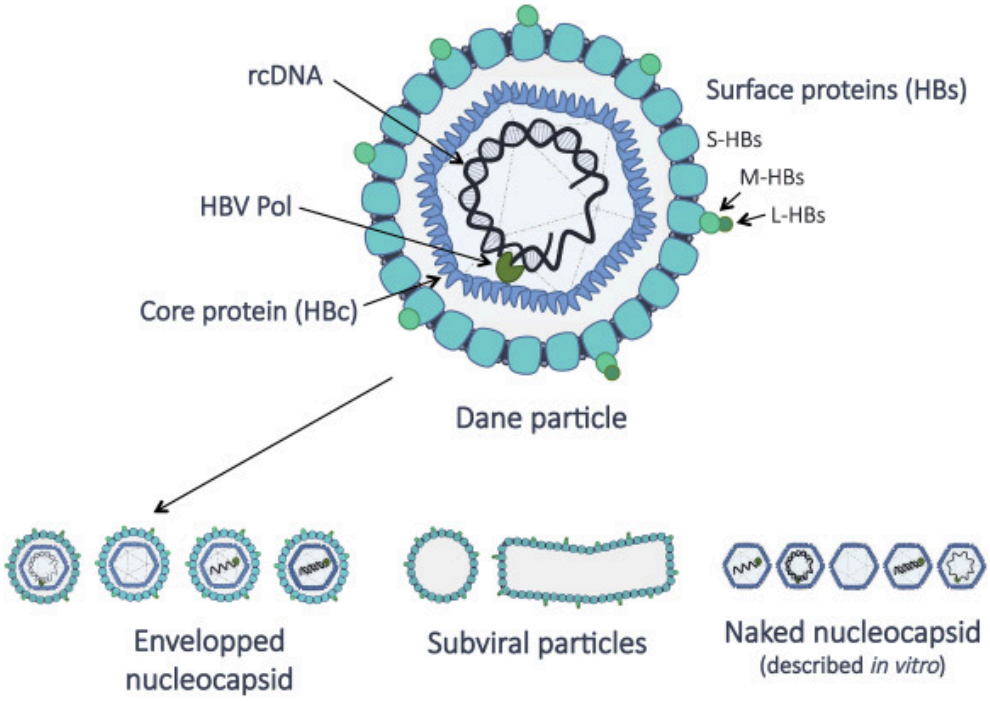


Figure 1.5. Schematic representation of HBV particles.

Infectious HBV virion (Dane particle) (upper) and non-infectious HBV particles, including enveloped capsids containing immature DNA/RNA, subviral particles (sphere and filament), and naked nucleocapsids (lower). Figure reproduced from[38] under a CC.BY license.

The HBV genome consists of a partially double-stranded DNA of approximately 3.2 kb (rcDNA: relaxed circular DNA) the negative DNA strand has a 'triple stranded' region, 'nick' and covalently bound polymerase and the positive strand is incomplete with variable length. rcDNA is converted into covalently closed circular DNA (cccDNA) in infected hepatocytes. cccDNA acts as a template for the production of HBV RNAs of different lengths which encode for HBV proteins (**Figure 1.6**).

The viral genome encodes four overlapping open reading frames (ORFs), C, P, S, and X, from which viral proteins are produced. The C ORF produces Hepatitis B core antigen (HBcAg), Hepatitis B e antigen (HBeAg) and pre-core protein, P produces Pol, S produces three types of surface antigens and X produces HBV X protein (**Figure 1.6**). Pol is the largest HBV protein, consisting of four domains 1) the terminal protein (TP) domain, which binds pregenomic RNA (pgRNA) and initiates negative strand DNA synthesis, 2) the spacer domain, 3) the reverse transcriptase (RT) domain, which has reverse transcriptase and DNA-dependent DNA polymerase activity and 4) the ribonuclease H (RNaseH) domain, which degrades pgRNA after reverse transcription. Mutations in the S gene, basal core promoter (BCP) and pre-core region of the C gene have been associated with cirrhosis and HCC[39,40].

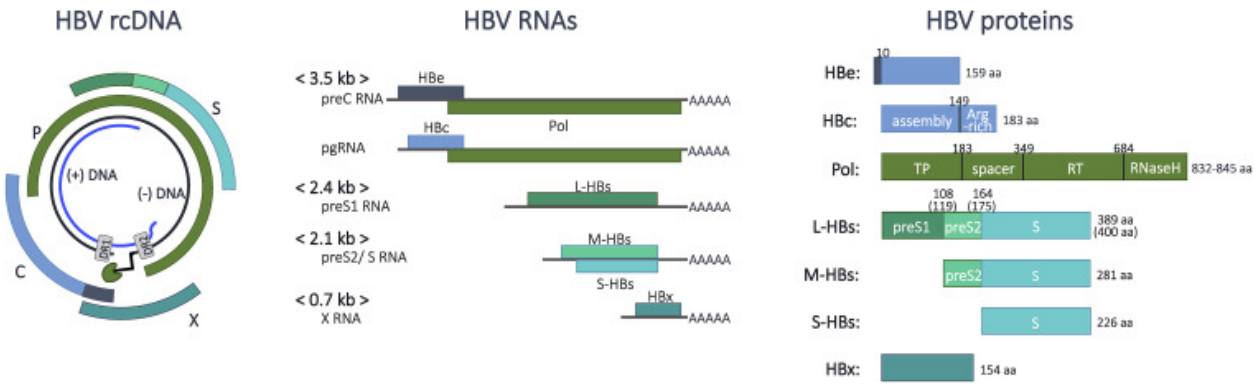


Figure 1.6. Schematic representation of the structure of HBV genomic DNA, RNAs, and proteins. Left, HBV genomic rcDNA and the encoded ORFs (C, P, S, X). Center, HBV RNAs (grey lines) produced by cccDNA transcription and the proteins (boxes) produced from the RNAs. The RNA lengths, as well as the names of the RNAs, are shown on the left. Right, HBV proteins and the domain structures. Amino acid numbers and the lengths are shown above the box and on the right, respectively. Reproduced from [38] under a CC.BY license.

1.3.5 HBV genetic diversity

HBV replicates via the reverse transcription of HBV RNA, since the viral polymerase lacks proofreading activity, mutations are common and as a result the HBV genome exhibits significant genetic diversity with an evolutionary rate of 10^{-5} - 10^{-6} substitutions per site per year[41,42]. There are ten recognized genotypes (A-J) with an inter-group divergence of >8% and multiple subgenotypes (**Figure 1.7**) which differ in their geographic distribution and clinical outcomes including response to treatment[43,44]. HBV genotypes are defined based on a phylogenetic tree of full genome sequences, the 8% threshold was based on the observed genetic variation between distinct clusters of sequences[45]. Accurate (sub)genotyping requires the analysis of full-length genome sequences to avoid recombinant strains being misclassified as novel subgenotypes. Different genotypes are thought to have originated in geographical locations[46,47], with migration impacting on the spread of genotypes globally. The same host can be co-infected by different HBV genotypes, which can lead to formation of recombinant virus. Recombination within and between genotypes generates novel variants,

including new genotypes (B2-B4 are recombinant viruses with the precore/core gene from genotype C) and plays a key role in current genetic diversity[48–50].

In the WHO African region, Genotype A is prevalent in southern, eastern, and central Africa, genotype D is dominant in northern Africa, and genotype E is most common in western Africa. While significant progress has been made in mapping the distribution of these genotypes across Africa, studies linking HBV genotypes to clinical outcomes remain limited. Genotype A has been associated with more severe clinical outcomes, including a higher risk of HCC in black Africans, compared to non-A genotypes[51]. Subgenotype A1 and E, found in southern and western Africa respectively, have been linked to higher risk of HCC. In Uganda, Genotype D and D/E recombinants were associated with higher viral loads (VLs) than genotype A[52].

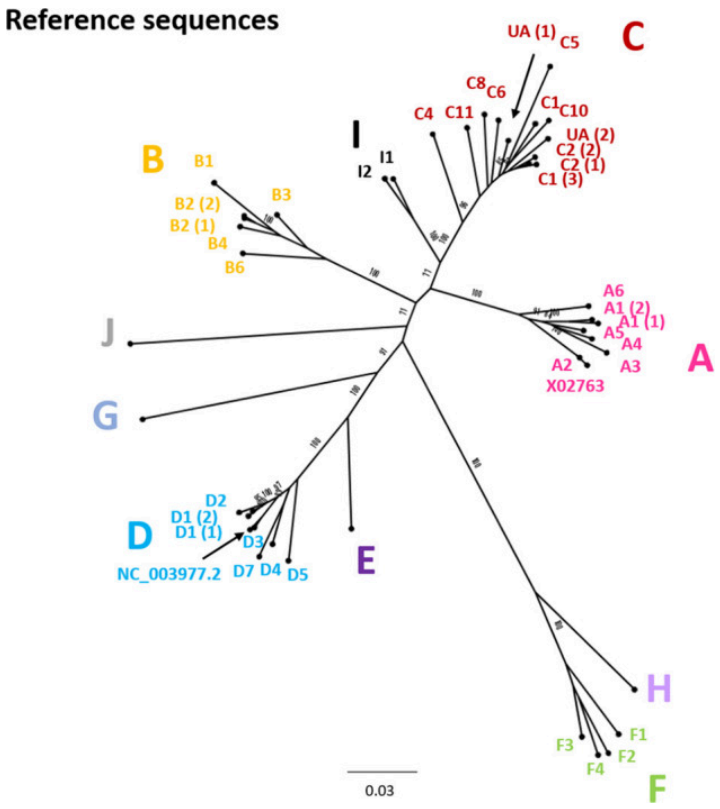


Figure 1.7. Genomic-length maximum-likelihood phylogenetic tree of HBV genotype, subgenotype and clade reference strains[53]. The genotype is given in each case and the subgenotype or clade identification is given where possible. Bootstrap support for branches ≥70 after 1000 replicates is indicated. The scale bar indicates the estimated nucleotide substitutions per site. Figure reproduced from McNaughton et al[53] under the terms of a CC.BY Licence.

1.3.6 HBV diagnosis and monitoring

Diagnosis of CHB requires the detection of HBsAg or HBV DNA on two occasions six months apart. HBeAg is associated with high levels of virus replication, individuals are typically more infectious if HBeAg positive, and less when HBeAg negative/anti-HBe positive. Occult HBV infection (OBI) is defined as the presence of replication-competent viral DNA in the liver tissue or blood of individuals testing negative for HBsAg. OBI may occur during the resolution of acute or chronic HBV infection where HBV DNA is still detectable after HBsAg loss (VL

usually <200 IU/ml), or due to HBsAg variants, which lead to diagnostic escape from HBsAg assays or affect HBsAg expression (VL similar to those with CHB)[54]. People with OBI are at risk of cirrhosis and HCC[55] and can lead to transmission via blood transfusion[56].

The field is moving away from old conventions of “disease phases”, however these have historically formed the basis of diagnosis, surveillance and treatment decisions. Monitoring of HBsAg, HBeAg, HBV DNA in addition to liver function (alanine aminotransferase - ALT and aspartate aminotransferase - AST), platelet count, transient elastography (TE) and ultrasound is now recommended to determine treatment eligibility and to monitor for complications such as liver fibrosis, cirrhosis and HCC[57,58]. Many areas lack affordable and accessible HBV DNA testing, a fundamental requirement for diagnosis, treatment and HBVDR surveillance. Liver biopsy, the traditional gold standard for assessing liver fibrosis, is invasive, costly, and impractical for routine use in many settings, therefore, non-invasive tests (NIT), such as AST-to-Platelet Ratio Index (APRI) and TE, are increasingly employed. TE, although effective (sensitivity ~80% for detecting F2 fibrosis[59]), remains inaccessible to most patients globally due to its high cost and limited availability[60]. WHO guidelines recommend using APRI, originally developed for use in chronic Hepatitis C infection[61], as the preferred NIT of liver disease.

1.3.7 HBV transmission

HBV is primarily transmitted through contact with infected bodily fluids, including blood, saliva, semen, and vaginal secretions[62]. The main modes of transmission are perinatal (from mother to child during childbirth), early childhood exposure to infected household contacts, unsafe medical practices (such as reuse of contaminated needles or instruments), unprotected sexual contact and injection drug use. Globally, perinatal transmission and early

childhood exposure are the predominant routes of transmission. Without intervention, the rate of vertical transmission from HBeAg-positive mothers is between 70-90%, and from HBeAg-negative mothers 10-40%[63–66]. The expansion of HBV vaccination programmes will change the global epidemiology over time, reducing the proportion of infections acquired at birth or during childhood.

1.3.8 HBV prevention strategies

The HBV subunit vaccine, based on the small HBsAg protein (which includes the 'a' determinant from a HBV A2 genotype, serotype *adw2),* was introduced in the 1980s, is safe and effective. There is homology between HBV genotypes in the 'a' determinant region, resulting in cross-protection across HBV genotypes, with the possible exception of the (rare) genotype F, where case reports exist suggesting that current vaccines may not confer full protection[67]. The established correlate of protection for hepatitis B vaccination is an anti-HBs antibody titre >10 IU/L. Vertical transmission can be prevented with interventions pre-natally (maternal HBsAg screening ± antiviral prophylaxis), peri-natally (birth dose (BD) vaccine ± hepatitis B immunoglobulin) and post-natally (3-4 follow-up vaccine doses, HepB3)[68]. This vaccine schedule leads to immunological protection in ~95% of children and is predicted to avert 38 million deaths among people born between 2000-2030 in 98 LMICs[69]. Universal vaccination of newborns has been particularly effective in reducing the incidence of HBV infection in countries with high endemicity. However, gaps in vaccine coverage continue to pose challenges to achieving HBV elimination targets. In 2000, BD coverage within 24 hours of birth was 5% globally, rising to 45% in 2022. This is set to improve further, as in 2023 Gavi, the Vaccine Alliance, recommitted to supporting BD vaccination, in addition to hexavalent infant vaccination which delivers HBV vaccine alongside other routine childhood immunisations in the first 2-6 months after birth[70].

Vaccination can also prevent horizontal transmission. Vaccinating high-risk groups, such as household members of individuals with CHB, HCWs, people who inject drugs, people who change sexual partners often, sex workers, people who have regular blood transfusions/products and men who have sex with men, reduces transmission[71]. These efforts, combined with harm reduction interventions, play a crucial role in lowering the global burden of HBV infection.

1.3.9 HBV treatment

For individuals with CHB, antiviral therapy plays a key role in reducing the risk of disease progression, improving long term survival and reducing transmission. The primary goal of treatment is to suppress viral replication, reducing liver inflammation, and preventing progression to cirrhosis and hepatocellular carcinoma.

Nucleos(t)ide analogues (NAs) are the mainstay of treatment, as they effectively suppress viral replication and reduce liver inflammation (**Table 1.1**). Historically interferon has been used, but this is often poorly tolerated and has a low rate of HBsAg clearance so the field has moved towards long-term oral NA therapy. Nucleoside analogues (e.g. entecavir (ETV)) and nucleotide analogues (e.g. tenofovir (TFV), administered as tenofovir disoproxil (TDF) or tenofovir alafenamide (TAF)) are first line treatments[72,73], while lamivudine (3TC) and adefovir (ADV) have been largely phased out due to the predictable selection of viral resistance-associated polymorphisms (RAMs)[74]. ETV and TFV are now widely used as first line agents, with regimen selection dependent on local availability and cost, evaluation of side-effect profiles (with ETV being preferred in individuals with renal dysfunction), comorbidity (with TFV being a natural choice to manage HIV/HBV coinfection), and exposure to prior NA therapy (with TFV preferred in this case).

A key challenge when evaluating HBV treatment response is the variation in definitions used to describe virological response. ‘Inadequate virological suppression’ can be considered as a definition which incorporates a spectrum of drug susceptibility and VL suppression dynamics. Definitions of ‘primary non-response’ and ‘partial response’ differ between existing guidelines (**Table 1.2**),[57,72,75] and there is no agreed distinction between persistent viraemia that is of potential clinical concern, versus a trajectory of declining VL on antiviral therapy, which is slow but nevertheless within a clinically expected range. Unlike 3TC, evidence suggests that partial response at 48 weeks on ETV may not indicate long-term failure, reducing the need for treatment adjustments. However, evidence is limited on when to implement "rescue therapy" for HBVDR. “Rescue” may involve either switching to a more potent agent with higher barrier to resistance and a non-overlapping resistance profile or to add a second agent without cross-resistance. Combination therapy may provide greater protection against multidrug resistance than sequential monotherapy and ETV and TFV may have additive antiviral activity when used in combination[76].

Table 1.1. Nucleos(t)ide analogue agents active against HBV. Adapted from WHO HBV treatment guidelines[75]

Nucleos(t)ide analogue (NAs)	Abbreviation	Class	Potency against HBV	Resistance barrier	Activity against HIV
Recommended agents					
Tenofovir*	TFV	Nucleotide	High	High	High
Entecavir	ETV	Nucleoside	High	High	Weak
No longer recommended					
Emtricitabine	FTC	Nucleoside	Moderate	Low	High
Telbivudine	LdT	Nucleoside	High	Low	Unclear
Lamivudine	3TC	Nucleoside	Moderate-high	Low	High
Adefovir	ADV	Nucleotide	Low	Moderate	None (at 10mg dose)

*TFV administered as TDF or TAF

Table 1.2. Comparison of international HBV guidelines for defining, monitoring and managing treatment failure.

World Health Organisation (WHO)[75,77]	European Association for the Study of the Liver (EASL)[57]	American Association for the Study of Liver Diseases (AASLD)[72]
Recommendations for monitoring		
VL testing available: Measure 3m in 1st year then every 6-12m. **VL testing not available:** suspect HBVDR based on NA received, low adherence, raised ALT, progressive disease.	Measure HBV VL every 3-4m during 1st year and every 6-12m thereafter.	Measure HBV VL every 3m until undetectable and every 3-6m thereafter.
Defining and managing primary non-response		
VL testing available: HBV VL not reduced by ≥1 $\log_{10}$ IU/mL within 3 months following initiation of therapy. **VL testing not available:** suspect HBVDR based on NA received, low adherence, raised ALT, progressive disease.	Defined at 3m. If good adherence, perform genotyping to help guide rescue strategy.	Defined at 6m in 2007 guidance, not included in 2018 guidance
Defining and managing secondary non-response (Virological Breakthrough, VBT)		
HBV VL rebound ≥1 $\log_{10}$ IU/mL from the nadir following an initial antiviral treatment effect (≥1 $\log_{10}$ IU/mL decrease in serum HBV VL)[78]. Recommend: switch agent	Adapt treatment as soon as identified/confirmed	Confirmatory testing and resistance testing. If VBT on ETV/TFV - switch or add 2nd agent
Defining and managing partial virological response		
If HBV VL not declining by 12m, switch or add another agent.	At 6m (3TC/LdT) or at 12m (ADV) - switch to ETV/TFV. At 12m (ETV/TFV) - if declining continue, if not add 2nd agent	At 24m for TFV/ETV . If persistent low level viraemia <2000 on ETV or TFV - continue monotherapy.
Specific recommendations about choice of antiviral agent		
3TC-, LdT- and ETV-resistant HBV can be treated with TFV monotherapy. ADV resistant infection can be treated with ETV or TFV. Insufficient responses to TFV may require ETV either as monotherapy or combination therapy.	Clear guidance based on prior NA exposure to switch/add ETV or TFV, avoiding sequential monotherapy with low barrier agents. For multidrug resistance, genotypic resistance testing is very useful, and a combination of a nucleoside and a nucleotide (preferably TFV) should be used.	Guidelines present clear recommendations for 'switch' and 'add' strategies for each agent. 3TC/ADV/ETV → TFV ADV/TFV → ETV
Guidance on resistance testing		
HBVDR may be confirmed through molecular testing. For multidrug resistance, genotypic resistance testing should be performed by a reference laboratory.	Ideally identification of the pattern of resistance mutations should be used to adapt therapeutic strategies.	Resistance testing can be useful in patients with past NA experience, those with persistent viremia on NA therapy, or those who experience VBT during treatment.

VL - viral load; 3TC - lamivudine, LdT - telbivudine, ADV - adefovir, ETV - entecavir, TFV - tenofovir, HBVDR - HBV drug resistance, m - months.

New WHO HBV treatment guidelines released in 2024[77], alongside national guidelines released by some countries (e.g. China[79], Brazil[80], and the Gambia[81]), update clinical and laboratory criteria for treatment and prophylaxis, to simplify assessment and expand eligibility to treatment (discussed further in **Chapter 6**). As therapy becomes more widely available, HBVDR should be considered as a potential barrier to its success, especially as most patients require long-term therapy to achieve sustained virological suppression. Persistent HBV replication due to HBVDR may have minimal clinical consequences, or may be associated with liver inflammation, long-term progression to chronic liver disease and/or hepatic decompensation[82–85]. Viraemia associated with HBVDR could also increase the risk of HBV transmission.

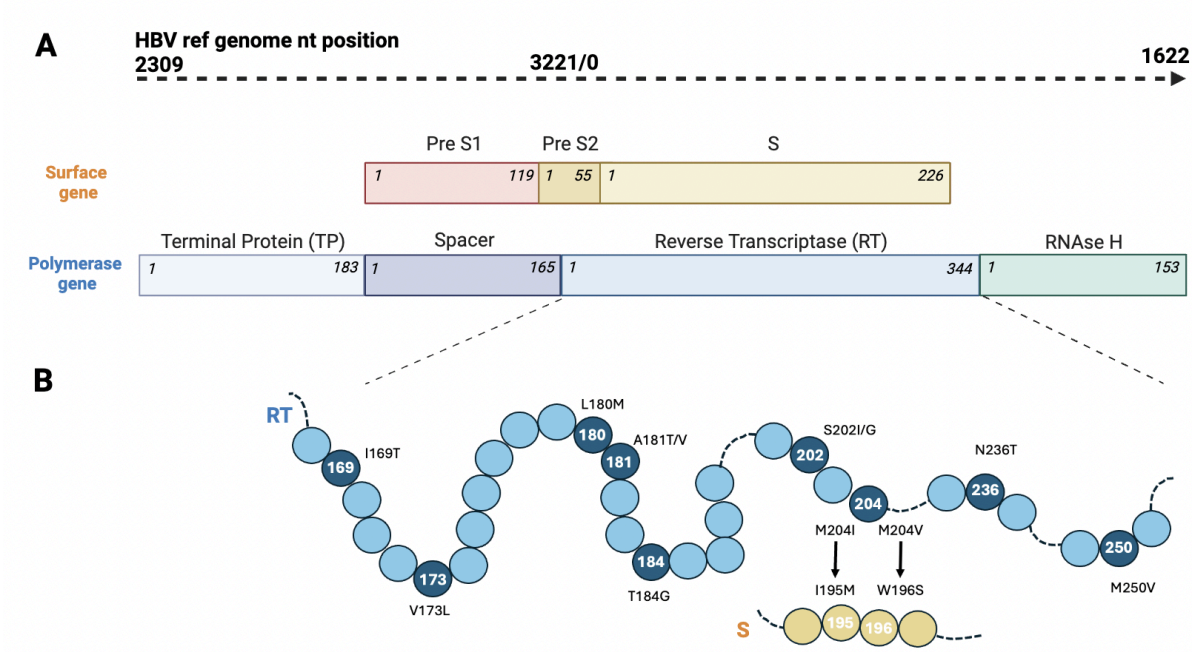
1.3.10 Mechanisms of HBV drug resistance (HBVDR)

HBVDR arises as a result of RAMs in the RT domain of the viral polymerase (**Figure 1.8**)[86,87]. While strong conserving restraints exist to maintain viral fitness and replicative capacity, exposure to selective pressures (including drug therapy) can drive the selection and persistence of viral polymorphisms. The best described RAMs arise in proximity to the active site of the viral RT, impeding binding of NA agents to their target[87]. The prototype RAM is a substitution of the wild-type residue methionine (M) with either isoleucine (I) or valine (V) at position 204 in RT, designated rtM204I/V, which is predictably selected by 3TC exposure and confers resistance to this agent. rtM204I/V also predisposes to ETV resistance; however, for an ETV-resistant phenotype to emerge, other RAMs must also be selected (a so-called ‘resistance haplotype’). Therefore, prior exposure to NA can influence subsequent development of drug resistance due to the presence of shared RAMs. For this reason, although ETV has a high genetic barrier to resistance, there is a specific risk in patients with

genotypic resistance to 3TC [88–91] with reported rates of ETV resistance as high as 30% in some studies[92,93].

TFV is used against HBV as a monotherapy, however is used as a combination therapy for HIV. HBV and HIV RT have only ~25% sequence homology, and equivalent sites which are associated with TFV resistance in HIV have not been associated with resistance in HBV[87]. Furthermore HBV may have a significantly higher 'genetic barrier' to resistance due to the overlapping reading frames in the HBV genome, RAMs in RT can also lead to changes in HBsAg encoded by the overlapping S gene, meaning RAMs may deleterious for the virus, potentially altering replication and transmission.

Changes in HBsAg encoded by the overlapping S gene, may also affect vaccine efficacy. These overlapping sites have been termed 'antiviral drug-associated potential vaccine escape mutants' (ADAP-VEMs)[94] **Figure 1.8**. Although not a recognised threat to the global HBV immunisation programme, ADAP-VEMs could have a public health impact if HBVDR becomes more widespread, highlighting another key rationale for surveillance.



C

HBV variant	NA drug agent				
	3TC	LdT	ETV	ADV	TFV
Wild-type	S	S	S	S	S
M204V	R	S	I	I	S
M204I	R	R	I	I	S
L180M + M204V	R	R	I	I	S
A181T/V	I	I	S	R	I
N236T	S	S	S	R	I
L180M + M204V/1 +/- I169T +/- V173L +/- M250V	R	R	R	S	S
L180M + M204V/I +/- T184G +/- S202I/G	R	R	R	S	S

Figure 1.8. Schematic to illustrate genetic basis of HBVDR. (A) Map showing the overlapping HBV polymerase and surface genes. Numbering is shown relative to HBVdb X02763. HBV polymerase (Pol) is divided into three functional domains and the spacer region. Pol contains between 832 or 845 amino acids (aa) depending on the genotype. Overlapping HBV surface antigen has three domains, PreS1, PreS2 and S. (B) Schematic view of the RT domain showing primary structure with amino acid residues in light blue, and sites of resistance associated mutations (RAMs) in dark blue with position number and relevant substitution indicated. Corresponding surface antigen mutations (antiviral drug associated potential vaccine escape mutations, ADAP-VEMs) are shown with amino acid residues in yellow. (C) RAMs alone and in combination that are reported in association with drug resistance based on EASL[57] guidelines; S = susceptible; I = intermediate/reduced susceptibility; R = resistant. 3TC - lamivudine, LdT - telbivudine, ETV - entecavir, ADV - adefovir, TFV - tenofovir. Figure from Lumley et al., under review at Lancet Microbe.

1.3.11 Definitions of HBVDR

Working with a group of WHO stakeholders as part of a WHO consultation on HBVDR (discussed further in **Chapter 5**), I assimilated working definitions pertaining to HBVDR and generated new proposed definitions, borrowing from HIV and/or Hepatitis C virus (HCV) terminology[95–98] (**Table 1.3).**

Table 1.3. Existing and proposed definitions pertaining to HBV drug resistance (HBVDR). Suggested definitions presented here are based on those commonly used in the literature and reviewed by an expert group.

Term	Definition(s)
Existing definitions	
Resistance associated mutation (RAM)	The amino acid substitution(s) in HBV RT that confer(s) resistance to drugs. To be robustly defined should be supported by evidence of drug resistance *in vitro* and *in vivo*.
Genotypic resistance [78]	Nucleotide substitutions lead to amino acid substitutions in HBV RT that confer resistance to antiviral drugs in phenotypic assays.
Definitions of HBV 'treatment failure' [75]	Definitions vary for primary and secondary treatment failure (VBT) and partial response- for more details see **Table 1.2**
Proposed definitions	
Acquired HBVDR (ADR)*	Emergence of HBVDR due to viral replication and selection of RAMs in individuals exposed to NA therapy. Relevant clinical history may include NA for PMTCT, period(s) of treatment interruption with or without documented virological failure, HIV treatment or PrEP containing HBV-active NA agents.
Transmitted HBVDR (TDR)*	HBVDR in NA-naive people due to primary acquisition of infection with a virus with one or more RAMs.
Pretreatment HBVDR (PDR)*	RAMs detected in HBV treatment naive people, or people with previous NA exposure(s), initiating or reinitiating first-line NA. May be ADR or TDR, or both.

NA – nucleos/tide analogue; PMTCT – prevention of mother to child transmission; PrEP – pre-exposure prophylaxis; RAM – resistance associated mutation(s); RT – reverse transcriptase; VBT – virologic breakthrough.

1.3.12 HBVDR epidemiology

The highest rate of acquired drug resistance (ADR) (definition in **Table 1.3**), is reported in association with 3TC exposure (~80% at 5 years), followed by ADV (29% at 5 years) and LdT (21% at 2 years in HBeAg positive infection)[96]. Although 3TC is no longer recommended as a first-line agent for HBV, 3TC-selected resistance remains a concern as it lowers the genetic barrier for resistance to ETV. Thus, while reported ETV resistance is low in NA-naive individuals (~1% by 5 years), this is higher in the context of prior NA exposure (~20% at 5 years), and in the presence of confirmed 3TC RAMs (>50% at 6 years), however no meta-analysis has been performed. The enhanced risk also applies even when RAMs cannot be detected,[91,99,100] suggesting that RAMs are novel or potentially archived in minority quasispecies, which exist below the threshold for detection in sequencing assays[101]. The prevalence of TFV resistance is deemed to be <1% at 10 years follow-up, but may vary between population settings and according to risk factors, including prior NA exposure[87,100,102,103]. There is limited evidence to document the prevalence of transmitted drug resistance (TDR), with current estimates varying widely by setting, therapeutic exposure and RAM, for example ranging from <1% to 37% in the WHO Africa region[104].

1.3.13 HBVDR testing

The current gold standard for HBVDR testing is sequencing (whole or partial HBV genome), to detect known and novel RAMs. Other approaches include line probe assays, real-time PCR, and restriction fragment length polymorphism (RFLP), focusing on the detection of known RAMs only[105,106]. There is not currently a unified approach to interpretation of an HBV genotypic resistance profile, in particular for TFV, as there is limited evidence regarding which mutations have a significant influence on susceptibility. Current bioinformatics tools are

limited, often lacking validation and consistent updates (**Figure 1.9**). One of the most widely used tools, geno2pheno does not list any TFV RAMs.

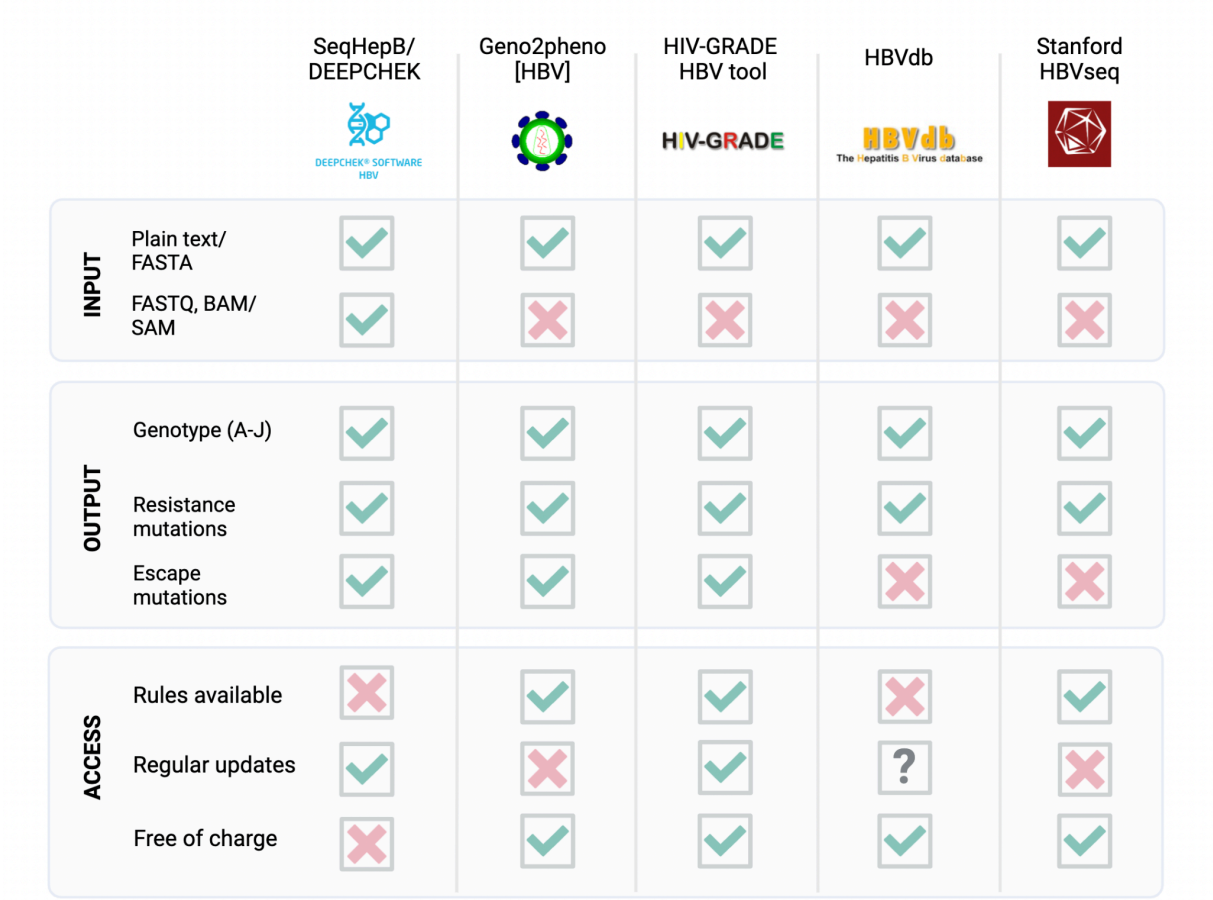


Figure 1.9. Comparison of current HBV resistance genotyping interpretation algorithms. Currently available analysis platforms vary in the accepted type of sequence input, mutations considered as significant, transparency, cost and update frequency. Created in BioRender. Lumley, S. (2025) https://BioRender.com/ m37p936

1.4. Viral whole genome sequencing as a translational tool

One of the themes cutting across my SARS-CoV-2 and HBV work is the use of viral WGS as a translational tool.

1.4.1 Applications of WGS to clinical and public health practice

WGS of viral pathogens has emerged as a powerful tool for understanding viral evolution, transmission dynamics, and resistance mechanisms[107]. For both SARS-CoV-2 and HBV, WGS has been instrumental in answering key clinical and public health questions. WGS has been critical in identifying new SARS-CoV-2 VOCs and informing vaccine and monoclonal antibody design and evaluation. For HBV, WGS facilitates the identification of different genotypes, subgenotypes and recombinant strains, which is important for predicting disease outcomes and guiding treatment decisions. In addition, WGS can identify HBVDR mutations in the viral polymerase gene, enabling personalised treatment strategies for patients with CHB and mutations leading to vaccine or diagnostic escape.

1.4.2 Shared sequencing challenges for SARS-CoV-2 and HBV

The two main challenges encountered when performing viral WGS, shared by HBV and SARS-CoV-2, are the presence of host background nucleic acid in clinical samples and low VLs of some samples, both of which can impact on genome coverage. A bi-modal distribution of VLs is seen in HBV, with a median VL of ~4 $\log_{10}$ IU/ml[108]; 50% of samples have VL below the current threshold required for generating full genome sequences (also ~4 $\log_{10}$ IU/ml) (**Figure 1.10**). These issues lead to difficulty obtaining full genomes using untargeted metagenomic sequencing and have led to the use of targeted viral sequencing approaches[107].

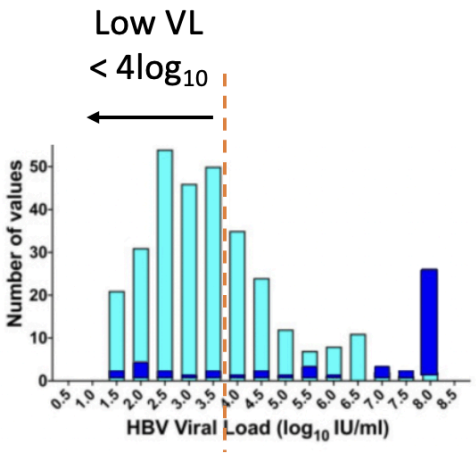


Figure 1.10. Bimodal distribution of HBV VL[108]. Orange dotted line shown at 4 $\log_{10}$ IU/ml, the median VL seen, below which I refer to as 'low VL'.

1.4.3 Unique challenges with HBV WGS

HBV poses additional challenges firstly due to its genetic diversity with >8% nucleotide divergence between genotypes. It is desirable to have sequencing methods capable of sequencing across all genotypes, however viral diversity makes the design of pan-genotypic capture probes and PCR primers challenging. Secondly the lack of a fully circular template in rcDNA leads to challenges with PCR primer design and amplification strategies such as RCA which rely on a fully circular template (**Figure 1.6**, discussed further in **Chapter 4**).

1.4.4 Approaches to targeted viral sequencing

The most common approaches for targeted viral sequencing are amplicon-based methods and hybrid capture techniques. Amplicon-based sequencing enriches the viral genome via PCR amplification using primers complementary to the virus of interest. Single or overlapping amplicons can be used to sequence whole or parts of viral genomes. Amplicon-based sequencing was used extensively during the pandemic for SARS-CoV-2 surveillance[109]. I use

an amplicon-based method in **Chapter 3** to establish in-house SARS-CoV-2 sequencing in the John Radcliffe hospital (OUH), and in **Chapter 4** I aim to capitalise on the rapid global expansion in expertise in amplicon sequencing by developing a similar sequencing method for HBV. Probe based capture uses biotinylated single-stranded DNA probes hybridise with target viral sequences within a sequencing library, and selectively capture target DNA, enriching the final library and increasing on-target sequencing yield, I use this approach for HBV in **Chapter 4**. I highlight the pros and cons of PCR amplification and hybrid capture in **Table 1.4**.

Table 1.4. Comparison of viral enrichment methods

	PCR amplification	**Hybrid capture**
Pros	Sensitive at low VL, straightforward design, well established methods, worldwide expertise	Overlapping probes increases tolerance for mismatches, greater % mismatch tolerated. Preservation of minor variant frequencies. Fewer PCR cycles (~ 20-25) limits introduction of amplification mutations
Cons	Primer mismatch in highly diverse pathogens, risk of lab contamination with amplicons, risk of amplification mutations with high number of PCR cycles (~ 35-40). Uneven amplification of PCR amplicons may impact minor variant calling.	High probe cost, lower sensitivity compared to PCR, low coverage in low VL samples.
Type of data	Read length dependent on amplicon length and sequencing platform.	Read length dependent on fragmentation/ sequencing platform used

PCR - polymerase chain reaction; VL - viral load

1.4.5 Approach to improving sensitivity of viral WGS methods

There are several stages shared between different target enrichment strategies that could be optimised to enhance sensitivity of viral WGS, which I use to improve the sensitivity of SARS-CoV-2 WGS in **Chapter 3** and HBV WGS in **Chapter 4**:

1. **Nucleic acid extraction:** efficient, high-quality extraction of viral nucleic acids is critical to generating libraries for sequencing. Using extraction kits specifically optimised for viral genomes, including the use of carrier material (e.g. carrier RNA in Qiagen virus spin kit) may improve yield in addition to optimising the input:output ratio to concentrate nucleic acids.
2. **Depletion of host nucleic acids**: in human samples, the majority of nucleic acids are of host origin, so selective depletion of human DNA may be useful alongside viral enrichment.
3. **Enrichment of viral nucleic acids**: optimisation of the targeted enrichment method e.g. optimising pre-amplification of viral genomes (e.g. with rolling circle amplification or PCR) or optimising targeted capture.
4. **Library preparation**: use of specialised library preparation kits designed for low-input DNA samples, minimising sample loss during library preparation, optimising bead clean ups of extracted material and adapting sample handling to target desired fragment size.

1.4.6 Sequencing platforms

I chose to develop protocols for Illumina and Nanopore platforms, as both are widely used for viral WGS, with expansion in global expertise and availability developed during the SARS-CoV-2 pandemic. Illumina uses a "sequencing by synthesis" approach in which DNA is fragmented and attached to a flow cell where each fragment is amplified to form clusters. A

DNA polymerase then adds fluorescently labelled nucleotides one at a time; the fluorescent signal indicates the identity of the nucleotide added, allowing the sequence to be determined. This method produces highly accurate short reads. Nanopore sequencing directly reads DNA molecules as they pass through a biological nanopore embedded in a membrane. As the nucleotides move through the nanopore, they cause changes in the electrical current. These changes are decoded into nucleotide sequences in real time, enabling generation of long reads which can easily capture a full HBV genome in a single read. In addition, ONT has multiple platforms e.g. MinION that are portable and can be operated independently of conventional laboratory facilities, and has been robustly and rapidly deployed in resource-limited settings as a point of care test[110].

The platforms differ in terms of strengths and limitations (**Table 1.5**):

Table 1.5. Comparison of strengths and limitations of Illumina and Nanopore sequencing platforms

Feature	Illumina	Nanopore (ONT)
Read Length	Short (100-600 bp)	Both short and long (>100 kbp)
Accuracy	High accuracy, low error rate: 97% Q30	Higher raw error rate (99.75% Q26, Q30 not reported)
Throughput	Platform dependent, from up to 8Tb data (Novaseq X)	Platform dependent, up to 9Tb data (PromethION 48)
Cost	Provider and sample number dependent, as little as £2/sample	Provider and sample number dependent
Portability	Non-portable	Portable option (e.g. MinION)

Bp - base pair; Q30 - Phred quality score of 30 representing an error rate of 1/1000bp, Q26 - quality score representing an error rate of ~1/400bp, Tb - terabyte.

1.5. Project rationale and thesis aims

1.5.1 SARS-CoV-2 projects and aims

During the first wave of the pandemic, it became apparent that HCWs in certain roles were at higher risk of infection, and infections persisted despite the use of PPE. I set up an in-house SARS-CoV-2 sequencing workflow at the John Radcliffe Hospital, Oxford, UK in November 2020, aiming to evaluate whether WGS could be used to enhance IPC practice by studying acquisition of SARS-CoV-2 in staff and patients with suspected nosocomial infections (**Chapter 3)**.

1.5.2 HBV projects and aims

My return to HBV research coincided with the preparation and release of the WHO 2024 treatment guidelines[74] and the updated Global Hepatitis Report[27]. Treatment expansion brings new challenges, for example, the potential for HBVDR. This raises critical questions about the risks of HBVDR in populations receiving broader access to antiviral therapies, the potential impact on global health and how these risks can be effectively monitored and mitigated.

In **Chapter 4**, I developed methods for HBV WGS, aiming to improve the sensitivity for HBV WGS and access to HBVDR testing. In **Chapter 5** I aim to (i) illustrate the interplay of factors that influence outcomes of NA therapy with a case report, (ii) summarise key concepts and roadmap for HBVDR based on a WHO multi-stakeholder consultation and my own literature review and (iii) summarise the risk of drug resistance to ETV and TFV in people receiving treatment for CHB with a systematic review and meta-analysis. Finally, in **Chapter 6** I aim to generate evidence to support the practical implementation of expanded WHO treatment

guidelines in resource-limited settings, exploring how these guidelines can be adapted to local contexts aiming to (i) evaluate treatment eligibility in a rural Ugandan setting and (ii) assess utility of different NIT for determining treatment eligibility. I then use WGS methods developed in **Chapter 4** across four settings in Africa, to characterise circulating genotypes, drug resistance and immune escape mutations, study household transmission and identify polymorphisms linked to clinical outcomes. This work aims to provide critical insights into the challenges of HBVDR and inform strategies to facilitate the success of global HBV treatment expansion.

Chapter 2: Methods

This chapter contains over-arching methods, including descriptions of cohorts, and approaches to development of wet laboratory sequencing protocols and bioinformatic pipelines. Individual results chapters contain more specific methodological details that are unique or newly developed in that chapter. In the work presented in this thesis, I first drew on samples collected in real-time in the context of the COVID-19 pandemic in Oxford, and then for HBV work I made use of samples collected through several pre-established research cohorts; the details of each of these is presented briefly below.

2.1. SARS-CoV-2 methods

2.1.1 SARS-CoV-2 setting and samples

Oxford University Hospitals NHS Foundation Trust (OUH) comprises four hospitals in South East England (three in close proximity in Oxford city and one approximately 45km away in the north of the county), with ~1100 beds (mostly in 4-bed bays within wards of 20-30 beds) and ~13,500 staff. In **chapter 3**, I refer to the four hospitals as “A”, a large acute hospital admitting both COVID-19 and non-COVID-19 patients; “B”, a smaller general hospital admitting both COVID-19 and non-COVID-19 patients; “C”, a hospital focused predominantly on cancer care; and “D”, a largely elective orthopedic hospital, with C and D not routinely admitting COVID-19 patients.

I obtained de-identified data from the Infections in Oxfordshire Research Database (IORD) which has generic Research Ethics Committee, Health Research Authority and Confidentiality Advisory Group approvals (19/SC/0403, 19/CAG/0144). The IORD project allows researchers access to routinely collected electronic data from the Oxford University

Hospitals NHS Foundation Trust Clinical Data Warehouse, to improve the management of infections (https://oxfordbrc.nihr.ac.uk/iord/; Chief Investigator Prof Sarah Walker).

SARS-CoV-2 PCR testing strategy

Infections in patients and hospital staff were detected by symptomatic and asymptomatic SARS-CoV-2 PCR testing of combined nasal and oropharyngeal swabs.

All patients received:

- Symptomatic testing as required;
- admission testing regardless of symptoms;
- weekly asymptomatic ward screening thereafter, increasing to twice weekly from 01-December-2020;
- for those who tested negative upon admission, a further single re-test was conducted between 5-7 days after admission;
- all patients were tested on discharge to other care settings, including to care homes or hospices;
- patient testing was conducted prior to admission for all elective patients.

All staff received:

- symptomatic testing as required;
- twice monthly clinic based asymptomatic staff testing programmes as previously described[19] ;
- voluntary twice weekly lateral flow device self-testing, introduced from 23-November-2020[111] ;

- lateral flow testing by the IPC team of all non-symptomatic staff (in addition to all patients and symptomatic staff) working in situations where there was an untoward incident or outbreak or high prevalence.

SARS-CoV-2 viral load quantification

SARS-CoV-2 VL quantification from combined nasal and oropharyngeal swabs of symptomatic and asymptomatic staff and patients was performed by the OUH clinical microbiology laboratory. Nucleic acid extraction was performed on the Kingfisher platform. Of the 2773 PCRs performed on this cohort during the study period, 2553 (92%) were performed using the Thermo Fisher TaqPath assay (targeting S and N genes, and ORF1ab; Thermo Fisher, Abingdon, UK), 78 on Altona RealStar (targeting E and S genes; Altona Diagnostics, Liverpool, UK), 64 on Abbott RealTime (targeting RdRp and N genes; Abbott, Maidenhead, UK), 59 on Cepheid Xpert® Xpress SARS-CoV-2 (targeting N2 and E; Cepheid, California, USA) and 19 on BioFire® Respiratory 2.1 (RP2.1) panel with SARS-CoV-2 (targeting ORF1ab and ORF8; Biofire diagnostics, Utah, USA). A result was considered positive on the Thermo Fisher platform if 2 out of 3 gene targets in addition to the positive control were detected (with Ct ≤32 for one target and ≤33 for another). Samples with all targets with Ct >37 or not detected were considered negative, other results were considered indeterminate. Extracts from PCR-positive samples were stored at -80°C for WGS.

2.1.2 SARS-CoV-2 laboratory methods

SARS-CoV-2 tiled amplicon primers

I evaluated two primer schemes; the ARTIC v3 hCoV-2019/nCoV-2019 amplicon set developed by the ARTIC project[112] which consists of 98 primer pairs generating 400bp amplicons and the Midnight v1 amplicon set designed by Freed et al.[113] which consists of 29

primer pairs generating 1200bp amplicons (see **Chapter 3**). Primer lists are available from the linked publications[112,113]. I took forwards the ARTIC v3 amplicon set for my SARS-CoV-2 sequencing work.

SARS-CoV-2 viral enrichment, library preparation and sequencing

I used the ARTIC "LoCost" nCoV-2019 sequencing protocol (LoCost) V.3 developed by Josh Quick et al (University of Birmingham, ARTIC network)[114] and sequenced libraries using R9.4.1 flow cells on the GridION (Oxford Nanopore Technologies, Oxford, UK). I performed the initial optimisation and sequencing work, Gillian Rodgers and Kevin Chau (Modernising Medical Microbiology group, University of Oxford) performed library preparation in the later stages of the project. Repeat isolates from the same individual were not routinely sequenced; if multiple samples per individual were stored, the lowest Ct sample (ie. highest VL) was selected for sequencing. I used the Qubit HS dsDNA kit (Q32854), as per the manufacturer's protocol, to determine the DNA concentration of SARS-CoV-2 PCR products and to quantify final sequencing libraries.

2.1.3 SARS-CoV-2 bioinformatic pipelines and data analysis

Consensus sequences were generated by Bede Constantinides (Modernising Medical Microbiology group, University of Oxford) using the ARTIC fieldbioinformatics v1.2.1 Nanopolish workflow[115]. All samples underwent quality control, requiring >90% consensus genome coverage at ≥20x depth. For 86 cases where duplicate samples were sequenced for an individual, the sample with highest percentage genome coverage was selected for analysis. A multiple sequence alignment (MSA) of 764 sequences based on Wuhan-Hu-1 was constructed using MAFFT parameters '--auto --6merpair --addfragments'. Phylogenetic reconstruction from the MSA was performed with IQTree parameters '-m GTR+G -blmin

1e-9'. The phylogeny was rooted with Wuhan-Hu-1 and branch lengths were SNP-scaled, phylogenies were prepared and visualised using Treeswift[116] and Toytree[117].

With supervision from Prof David Eyre, I performed combined epidemiological and genomic analysis of SARS-CoV-2 data using R version 4.0.2[118], with visualisation using ggplot2[119] and igraph[120] packages.

2.2. Hepatitis B virus methods

2.2.1 HBV setting and samples

Mock HBV samples

In addition to drawing on HBV clinical samples, I used mock samples to assess the performance of my sequencing workflows across all HBV genotypes, including rarer genotypes F-J that were unlikely to be present in our cohorts. Including these rarer genotypes is important, as they are often missing in reports of sequencing schemes however may become increasingly relevant as more HBV WGS is performed from diverse global regions. I used plasmids containing ~1.3 copies of the HBV genome representing genotypes A-J in a pUC57 backbone (**Figure 2.1**). Genotypes A-F were designed by Peter Revill[121], genotypes G-J were designed by Marion Delphin (Francis Crick Institute). I prepared mock HBV samples by spiking these plasmids into non-infected human serum (Merck, H5667) to simulate the human nucleic acid background in a sample, at a concentration of 10^5 plasmid copies/ml.

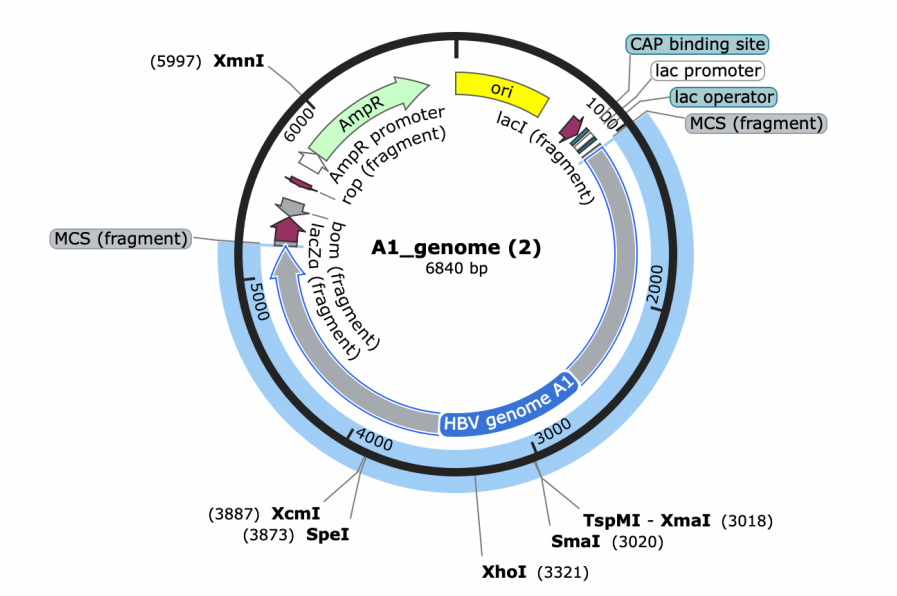


Figure 2.1. Example HBV plasmid construct with a 1.3mer HBV sequence inserted into the vector pUC57. HBV insert for subgenotype A1 shown as an example[121].

HBV Cohort descriptions and ethical approvals

I used blood samples from adults living with CHB recruited through cohorts in the UK, South Africa, Uganda, DRC and Kenya (**Table 2.1**). Adults were approached with information about the study and valid informed consent was collected by a member of the study team, prior to collecting demographic and clinical data, and blood samples. All cohorts except the UK blood bank cohort consisted of a mix of HBV mono-infected and HBV/HIV co-infected individuals. A full sample list is given in **Appendix 2**.

The clinical team at each location collected venous blood samples in EDTA tubes. To separate plasma, they centrifuged whole blood at 1800 rpm for 10 minutes, harvested the supernatant and stored it at −80°C. In addition, dried blood spots were collected by the clinical team in the DRC cohort, these were dried, and stored at room temperature.

Table 2.1. Description and ethical approvals for HBV cohorts in the United Kingdom and Africa

Cohort	Description	Ethical approval
UNITED KINGDOM		
UK, Oxford (OXSAHEP)	624 individuals from outpatient hepatitis clinic in Oxford[108,122]	Oxford Research Ethics Committee A (ref. 09/H0604/20) (PIs Philippa Matthews, Eleanor Barnes)
UK (UCL)	Outpatient hepatitis clinic in London	Characterising and modifying immune responses in chronic viral hepatitis, REC reference: 11/LO/0421 (PI William Rosenberg)
UK (Blood bank)	Blood donors with VL <4 log 10 IU/ml, 23 samples: 19 HBsAg positive, 4 HBsAg negative (OBI), genotypes A to E.	Blood Supply Clinical Audit, Risk and Effectiveness Committee of NHS Blood and Transplant on 19/04/22 (22031) and 08/01/23 (23002) (PI Heli Harvala)
AFRICA		
South Africa (OXSAHEP)	Outpatient clinics in Tygerberg Hospital, Cape Town (Stellenbosch University) and public sector hospital services in Bloemfontein (University of the Free State), South Africa[108]	University of the Free State (ref. UFS-HSD2018/0193-0001), OxTREC (ref 1-18) (PIs Philippa Matthews, Tongai Maponga Dominique Goedhals)
Uganda (ULiDs)	Uganda Liver Disease Study: community based longitudinal study nested in Uganda General Population Cohort (GPC), Kalungu District, south west Uganda[123,124]. Individuals seropositive for HBsAg were offered HBV VL, FBC, LFTs, TE and HBV WGS.	Uganda Virus Research Institute Research and Ethics Committee (ref. GC/127/12/11/06 and GC/127/19/04/711), Ugandan National Council for Science and Technology (HS870), OxTREC (ref 50-18). (PIs Philippa Matthews, Rob Newton)
Democratic republic of Congo (HOVER-HBV)	Cross-sectional study of participants from households of index HBsAg positive mothers initially identified through antenatal HBV screening in Kinshasa [125]. Established by collaborators Professor Thompson and Professor Parr (University of North Carolina)	Institutional Review Board at the University of North Carolina (19–1875) Ethics Committee at Université Protestante au Congo (CE/UPC/0062). (PI Jonathan Parr)
Kenya (STRIKE-HBV)	Cross-sectional study of 105 individuals recruited from outpatient clinic, Kilifi County Hospital (non-pregnant) and screening of close contacts.[126] Individuals seropositive for HBsAg were offered a comprehensive liver health assessment including TE, HBV VL and HBV WGS. Established by Dr Louise Downs.	Kenya Scientific and Ethics Review Unity (SERU) (number 4656), OxTREC (ref 22-23). (PI Louise Downs)

OBI - Occult HBV infection; TE - transient elastography; VL - viral load; WGS - whole genome sequencing

HBV viral load quantification

HBV VL was determined using one of a range of assays detailed in **Table 2.2**.

Table 2.2. Assays used for HBV VL quantification across settings

Assay	Lower limit of quantification	Cohort and testing laboratory
Abbott M2000 HBV DNA assay	20 IU/ml	United Kingdom (Oxford) and Kenya @ Oxford University Hospitals (OUH) diagnostic microbiology laboratory, UK
		Democratic Republic of the Congo @ University of North Carolina, USA
Cobas Ampliprep/Taqman HBV test	10 IU/ml	South Africa @ National Health Laboratory Services, Tygerberg Hospital, South Africa
		Uganda @ Medical Research Rouncil/Uganda Virology Research Institute/London School of Hygiene and Tropical Medicine Clinical Diagnostic Laboratories, Uganda
Aptima® HBV Quant Assay	6 IU/ml	United Kingdom (UCL) @ University College Hospitals NHS Foundation trust microbiology laboratory
qPCR[127]	1 IU/ml	United Kingdom (Blood bank) @ Oxford University (by Michael Fu)

2.2.2 HBV laboratory methods

Host depletion

In order to remove as much background host material as possible (thus enriching for our HBV target), I compared two different approaches for host depletion (see **Chapter 4, section 4.4.2** for comparison): (1) I depleted host nucleic acid by adding 1.25µl micrococcal nuclease (NEB catalogue number M0247S) and 23.75µl mnase buffer to 250µl plasma and incubated for 20 minutes at 37°C, using a protocol developed by Delphine Bousquet (Université de Lyon)[128], (2) I performed host depletion using the NEBnext microbiome kit (NEB catalogue number E2612) as per manufacturer's instructions.

Nucleic acid extraction

I used a variety of methods for nucleic acid extraction, described in **Table 2.3**.

Table 2.3. Methods for nucleic acid extraction from samples obtained from individuals with CHB across settings.

Project	Extraction kit	Input/output volume
Takara-capture (section 4.5.2, figure 4.7)	NUCLISENS easyMag total nucleic acid extraction system (bioMerieux)	1ml plasma eluted into 50 µL of kit elution buffer
Rolling circle amplification (section 4.4) and NEB-capture (section 4.5.2, figure 4.7)	Kingfisher Apex Magmax viral/pathogen nucleic acid isolation kit (cat no. A42352)	2x 250µL aliquots of mnase-treated plasma eluted into 50µl of kit elution buffer with SPRIselect bead clean up to reduce volume to 40µl
HEP-TILE (Oxford, Kenya*, Uganda) (section 4.6 and 6.5)	QIAgen MinElute Virus Spin kit (cat no. 57704) with carrier RNA	200-400µL plasma eluted into 45µL nuclease free water
HEP-TILE (DRC)** (section 4.9 and 6.5)	Venous blood: QIAgen DNA blood midi kit spin protocol (cat no. 51104)	2ml sample eluted into 110µL elution buffer with eluent run twice through the column.
	Dried blood spot: Chelex extraction[129]	Single 6mm DBS punch, eluted into 150µL
HEP-TILE (NHSBT low VL) *** (section 4.8)	Roche High Pure viral nucleic acid large volume extraction kit (cat no. 05114403001)[127]	5ml plasma extracted in 50µL kit elution buffer

DRC – Democratic Republic of Congo; NHSBT – NHS Blood and Transplant; DBS - dried blood spot

* Extractions performed at KEMRI-Wellcome Trust Research Programme (KWTRP) by Dr Louise Downs

**Extractions performed by collaborator Camille Morgan, University of North Carolina

***Extractions performed by collaborator Michael Fu, University of Oxford

HBV rolling circle amplification (RCA)

For the original RCA protocol I used RCA primers designed by Margeridon et al.[130] (**Table 2.4**). For the 'gPCR-RCA' protocol (described in **Chapter 4**), I used primers from Gunther et al[131] to generate full length HBV amplicons with a SapI restriction site at each end (**Table 2.5**), followed by digestion with SapI and ligation to generate double stranded circular HBV DNA without nonhomologous sequences.

Table 2.4. Rolling circle amplification primers from Margeridon et al.[130]

Primer	Primer Sequence	Nucleotide position
RCA1	AATCCTCACAATA*C*C	226–240
RCA2	GATGGGATGGGAA*T*A	615–601
RCA3	CCTATGGGAGTGG*G*C	637–651
RCA4	GCAACGGGGTAAA*G*G	1154–1140
RCA5	ATGCAACTTTTTC*A*C	1814–1828
RCA6	TCCAAATTCTTTA*T*A	1930–1916
RCA7	TAGAAGAAGAACT*C*C	2368–2382
RCA8	AGAATATGGTGAC*C*C	2828–2814

*Denotes position of a phosphorothioate modification to counteract 3´→5' proofreading exonuclease activity).

Table 2.5 Full length gPCR primers from Gunther et al.[131]**.** Restriction sites underlined.

Primer	Primer Sequence	Nucleotide position
P1 (S)	CCG GAA AGC TTG AGC TCT TCT TTT TCA CCT CTG CCT AAT CA	1821-1841
P2 (AS)	CCG GAA AGC TTG AGC TCT TCA AAA AGT TGC ATG GTG CTG G	1825-1906

HBV probe design

HBV probes were designed by Dr Azim Ansari (University of Oxford), the probe sequences and the set of HBV genome sequences that were used for their design can be downloaded here[132]. As previously described[133], a maximum likelihood phylogeny of HBV full genomes was used to infer the ancestral root sequence which was divided into 120 nucleotide segments with 60 base overlaps which resulted in an initial set of 55 probes (**Figure 2.2**). Additional probes were added to cover the insertion/deletions found in genotype G and D respectively and to span the break point '0' used to display the genome linearly. As previous work in HCV probe-based sequence capture had demonstrated that probes of 120 bases long can tolerate up to 20% divergence relative to their target sequence before the efficiency of capture drops[134], for each viral sequence, if a continuous region of ≥60 bases diverged from the probe sequences by >20%, a new probe was designed for the region using the genotype consensus sequence. The final probe set contained 74 probes. The probe sequences and the set of HBV genome sequences that were used for their design can be downloaded using this link: https://figshare.com/articles/dataset/HBV_probe_sequences/22127015

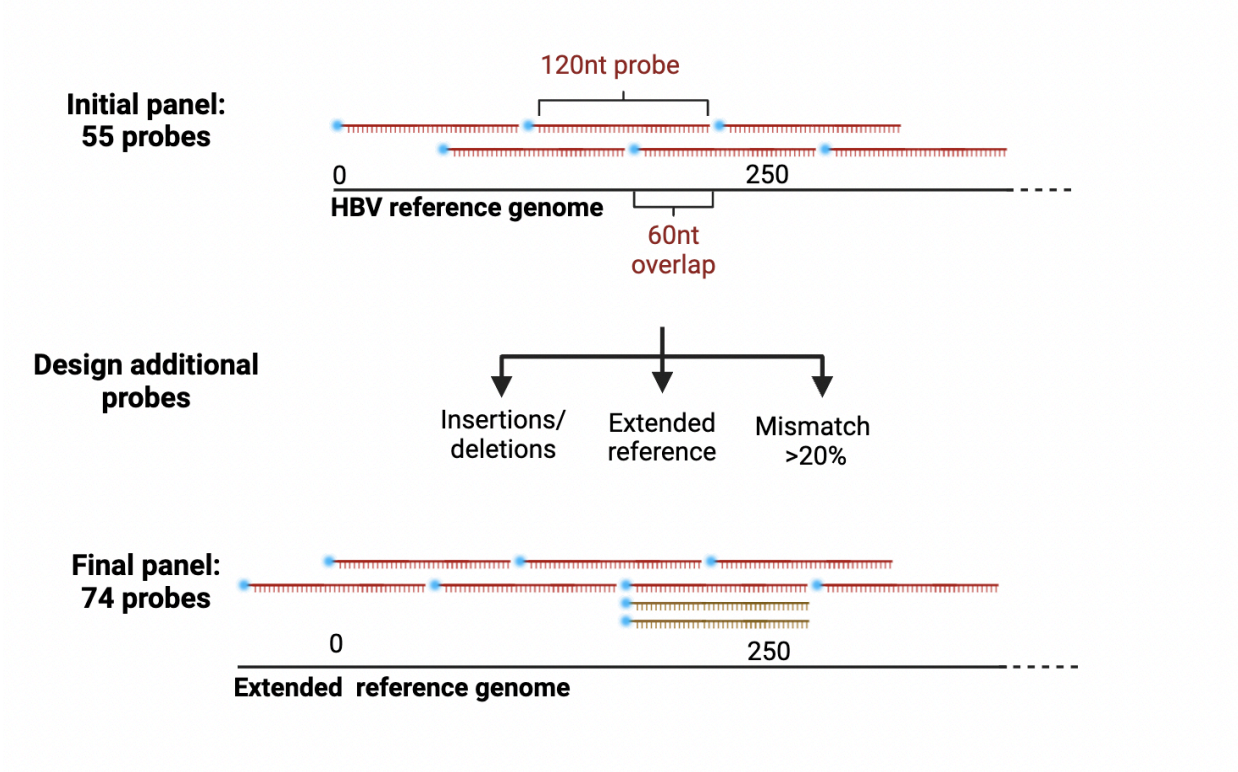


Figure 2.2. Schematic to show the process of probe design for HBV probe-based capture. An initial panel of probes (red) was designed by dividing the ancestral root sequence into 120nt segments with 60nt overlap. Additional probes (brown) were added to cover the insertion/deletion found in genotype G and D respectively, to cover the break point '0' used to display the genome linearly and if a viral sequence diverged from the probe sequences by >20%. The final probe set contained 74 probes. Blue dots indicate 5' biotinylation of probes.

HEP-TILE primer design

In collaboration with Chris Kent (University of Birmingham) I developed a pan-genotypic HBV scheme using an early version of PrimalScheme3 (https://github.com/ChrisgKent/primalscheme3), the successor to PrimalScheme[135], a web-based primer design tool for developing multiplex primer schemes. Chris made a number of changes in PrimalScheme3 to enable us to generate a tiled amplicon scheme which covered the circular HBV genome, utilising a number of discrete primers at each position to handle intraspecies diversity.

Amplicon schemes generated by the original PrimalScheme[135] resulted in a short sequence at each end of the reference genome which is not covered due to primer placement constraints and downstream primer-trimming. Since the conventionally designated 'start site' for the circular HBV genome falls within the overlapping genes encoding surface and reverse transcriptase, this results in the loss of epidemiologically and clinically relevant information. PrimalScheme3 introduces an option to add an amplicon spanning the start/end of circular genomes, resulting in 100% genome coverage (**Figure 2.3**).

Intra-species diversity poses an issue for amplicon sequencing, as variation within the primer binding sites reduces primer binding efficacy. The original PrimalScheme identified conserved primer binding sites by heavily penalising those with variation within the input genomes. This approach is effective for closely related genomes, such as outbreak strains, but for diverse inputs a different approach is needed. PrimalScheme3 handles diversity with the use of 'primer clouds' i.e. a discrete set of 3'-anchored primers which cover all mutations above a user-specified frequency, without the use of ambiguous bases. This approach reduces the negative effects of adding additional primers to a multiplexed PCR, such as primer-primer interactions, and mispriming.

Primer design was based on a download of all available HBV non-recombinant whole genomes from the Hepatitis B Virus Database (HBVdb)[136] on 15th February 2024, in addition to genomes for genotypes I and J sourced from McNaughton et. al[53]. Chris aligned HBVdb HBV genomes from each genotype (A-H) separately using MAFFT[137], and phylogenetically downsampled each genotype to 0.95 relative tree length using TreeMMer[138]. He combined each downsampled genotype's genomes with genotype I and J sequences from[53], aligned with MAFFT[137], and input into PrimalScheme3 command line tool.

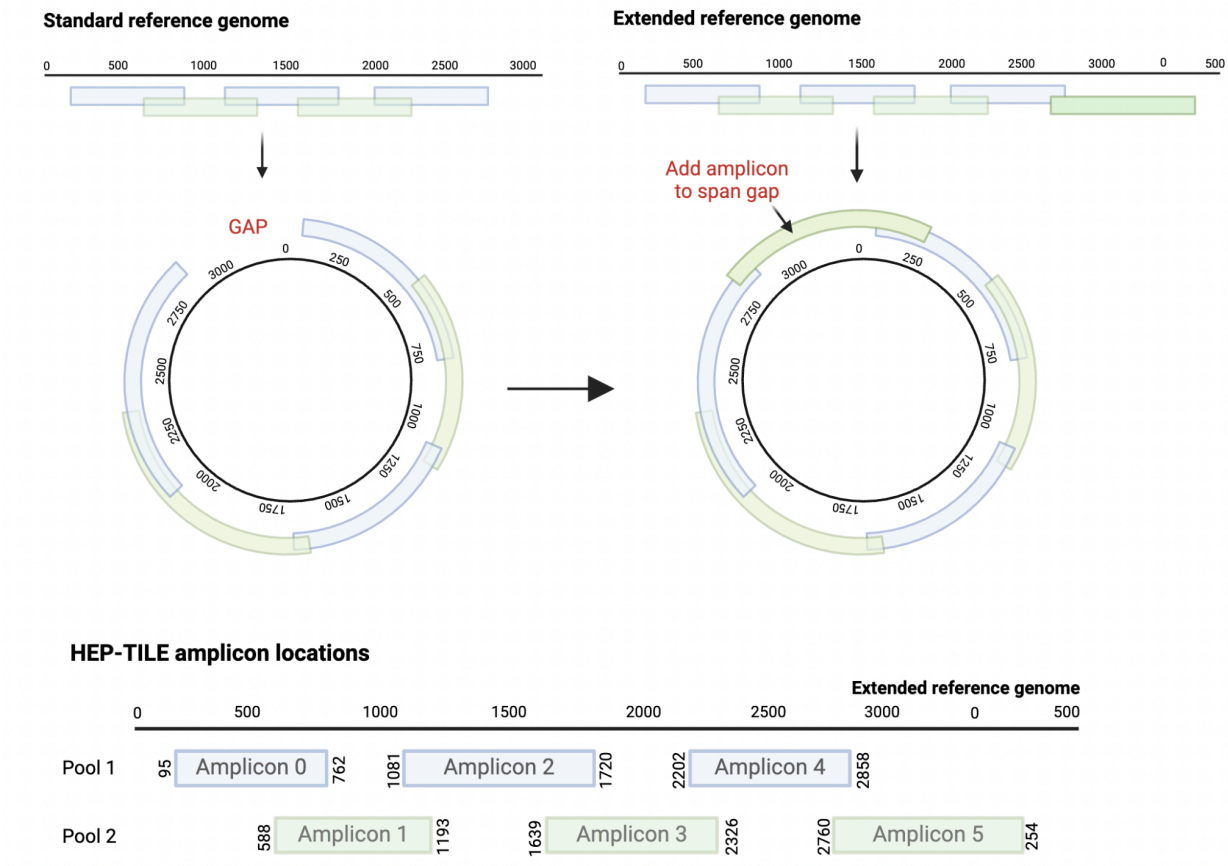


Figure 2.3. Schematic to show the design of multiplex tiled PCR for the circular HBV genome. Generation of tiled schemes from linear reference genomes leaves a gap in coverage for a circular genome. An additional amplicon spanning the start/end of the linear reference ensures a theoretical maximum 100% genome coverage can be obtained. Primer positions and amplicon length for HBV tiled amplicon scheme. Numbering of nt positions based on X02763 reference genome. Figure from Lumley et al., under review Scientific Reports.

Initially, we developed a 500bp, eight amplicon scheme[139], which produces amplicons that can be sequenced on both Nanopore and Illumina without fragmentation. However when trialled, one amplicon spanning 1700-2000 nt region of the HBV genome consistently dropped out in samples with VL <5 $\log_{10}$ IU/ml (**Figure 2.4 A and B)**. I hypothesised that this could be explained by the structure of HBV DNA: in low VL samples only rcDNA is present, which has discontinuities in both the positive and negative DNA strands in that region leading to amplicon dropout. In contrast, in higher VL samples, covalently closed circular DNA (cccDNA) is also present in the plasma[140,141]. cccDNA has continuous positive and negative strands which act as a template for this amplicon. I identified a suitable location for primers spanning this region, taking into account the structure of rc-DNA, and re-designed a 600bp primer scheme (hbv/600/v2.1.0)[142] (**Figure 2.4 C-E**), resulting in a six amplicon scheme with amplicons ~600-715bp long, generating a theoretical 100% HBV genome coverage (**Figure 2.4 F**). This scheme uses 131 primers to cover the HBV diversity present in our reference dataset.

Primer sequences are available open access:
https://github.com/quick-lab/primerschemes/tree/main/primerschemes/hbv.

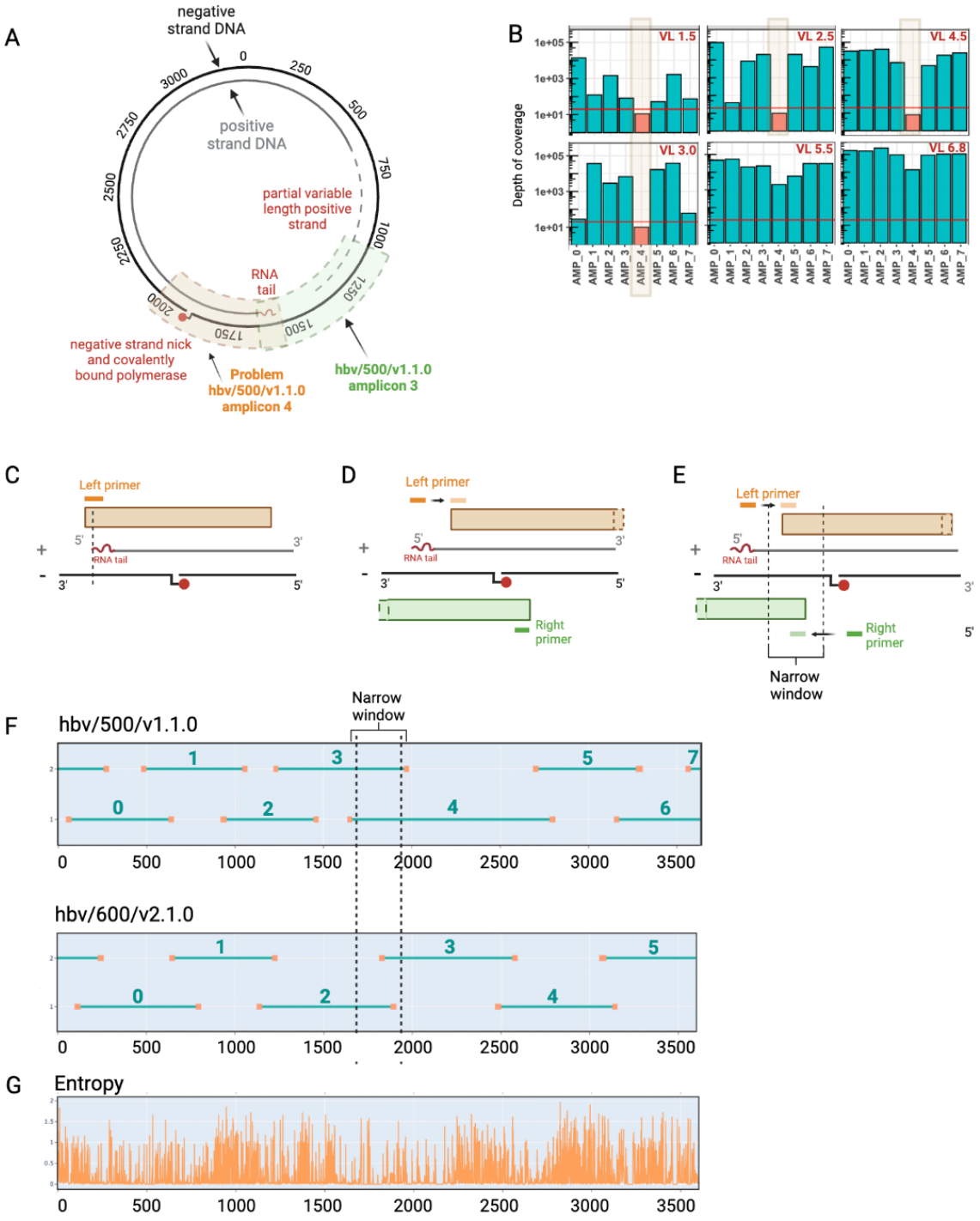


Figure 2.4. Modifications to HEP-TILE primer positions to take into account HBV genome structure. (A) Schematic of HBV relaxed circular DNA (rc-DNA), the HBV partially dsDNA genome has 5' modifications to both the positive DNA strand (RNA tail) and negative DNA strand

(covalently bound polymerase). The negative strand is complete but not fully circular due to the presence of a ‘nick’ and covalently bound polymerase. The partial positive strand is variable in length. (B) Depth of coverage for amplicon scheme hbv/500/v1.1.0 - illustrating drop-out of the ‘problem amplicon’ (initially Amplicon 4) in samples below a VL of 5 $\log_{10}$ IU/ml (VL shown in red). (C) Amplicon B drops out in low-medium VL samples, the hypothesis is that neither the positive nor negative strands of the HBV relaxed circular genome are complete in that region. The left hand primer only partially overlaps the 5’ tail of the positive strand, the negative strand contains a nick. (D) Attempt to solve issue by moving the amplicon B primer upstream, however this has the knock on effect of the Amplion A right hand primer moving upstream, which now means amplicon A falls across a non-continuous DNA region. (E) Solution requires the left hand primer for amplicon B and right hand primer for amplicon A to be placed in a narrow 200bp window, to ensure Amplicon B can be generated from the positive DNA strand, and amplicon A from the negative DNA strand. (F) Locations and numbering of primers for initial 8 amplicon primer scheme (hbv/500/v1.1.0) and updated 5 amplicon primer scheme (hbv/600/v2.1.0) Created in BioRender. (G) Entropy across the HBV genome. Lumley, S. (2024) BioRender.com/l10u171

Viral enrichment and library preparation protocols

In **Chapter 4** I describe the optimisation of three HBV enrichment techniques; rolling circle amplification (RCA), probe based capture and tiled amplicon (HEP-TILE). For each technique I trialled a range of different optimisations and library preparation protocols, detailed in **Table 2.6** below, and discussed in detail in **Chapter 4**.

Table 2.6. Methods for HBV enrichment and library preparation

Method	Protocol details and references	References
HBV rolling circle amplification		
CL/RCA (section 4.4)	**Phi29 protocol**[143] **Equiphi protocol** - Substitute Phi29 buffer/enzyme for EquiPhi29 DNA buffer/polymerase - Incubate for 45C for 2 hours, inactivate at 65C for 10min. **Debranching**: - Dilute RCA products two-fold with nuclease free water - Purify RCA products using 0.6x SPRI beads following manufacturers recommendations - Prepare debranching reaction using 1.5uL T7 endonuclease I, 3ul NEBuffer 2 10X, purified RCA product (variable ul) made up to 30ul with nuclease free water. Incubate for 1 hour at 37C.	Phi 29[143] (adapted from[130,144])
FEN1-CL/RCA (section 4.4)	- Prepare 8.5ul HBV DNA, 2ul rCustSmart Buffer (B6004S), 2ul dNTPs (10mM), 1ul Equiphi DNA polymerase (A29290), 1ul Hi-T4 DNA ligase (M2622), 1ul FEN1 (M0645S), 0.8ul DTT (0.1M), 2ul ATP (10mM) made up to 20ul with water - Incubate at 47C for 50mins, inactivate at 65C for 10mins	
gPCR- RCA (section 4.4)	- Full length PCR: 7.6ul water, 0.4ul dNTP (10mM each base), 0.6ul P1 (10uM), 0.6ul P2 (10uM), 0.3ul Expand Hifi enzyme, 2ul buffer (10x conc with MgCl2) and 8.5ul template DNA (20ul total reaction volume). - Initial denaturation 94C for 2 minus, 40 cycle pcr with denaturation at 94C for 40s, annealing 60C for 1.5 min (60C for 5 cycles, 66C for 5 cycles, 72C for 30 cycles) and elongation at 70C for 3min for 10 cycles then 5min for 30 cycles. Final elongation 72C for 7 minutes. - Run PCR product on 1.2% TAE agarose gel at 100V for 40mins. Visualise under UV light and cut out 3.2kb band, weigh band and store in eppendorf at 4C. - Purify with Monarch T1020S DNA gel extraction kit and qubit. - Cut with restriction enzyme: 15ul DNA, 4ul 10X rCutSmart Buffer, 0.1ul SapI made up to 40ul with water. Incubate at 37C for 15mins, heat inactivate at 65C for 20mins. - Circularise using 2ul T4 DNA ligase, 5ul T4 ligase buffer (10X), 40ul DNA from step above and 3ul water, incubate at room temperature for 10mins,	Full length PCR [131] RCA[143]

	heat inactivate at 65C for 10mins. - Remove linear material prior to RCA using 0.5ul exonuclease V (M0345), 5ul NEBuffer 4(10X), 5ul ATP (10mM), 16ul DNA from previous step and 23ul H20. Incubate at 37C for 30min, inactiavate at 70C for 30min. - SPRI bead clean up and elute in 16ul nuclease free water - Phi29 RCA	
uDumbell (section 4.4)	- Protocol adapted from Limberis et al.	Limberis et al [145]
HBV probe based capture		
Takara library prep (section 4.5)	- Performed by Oxford Genomic Consortium - Takara Bio SMARTer Stranded Total RNA-Seq Kit v3 - Pico Input Mammalian (cat no. 634485) - OGC's in-house indexed primers - As per manufacturer's recommendations with sample volumes concentrated to 3µl using SPRIselect beads (Beckman Coulter) and processed using a 0.25x reduced scale miniaturised reaction volume for first strand cDNA synthesis and 12 indexing PCR cycles.	Lumley et al[133]
NEB library prep (section 4.5 and 4.7)	- NEBNext Ultra II FS DNA (NEB #E7805S) library preparation kit (protocol for use with inputs <100ng) - NEBNext Multiplex Oligos for Illumina (Unique Dual Index UMI Adaptors DNA Set 1) (NEB #E7395) - As per manufacturer's protocol with 26ul DNA input, 3 minutes fragmentation and 13 indexing PCR cycles.	Lumley et al[146] NEB Ultra II FS protcol[147]
IDT probe-based capture (section 4.5 and 4.7)	- Same method used by OGC (for Takara-capture library) and myself (NEB-capture library) - IDT AMPure XP Bead DNA concentration and tube protocols - 4.4µg aliquot of the final multiplexed library concentrated then enriched for HBV using the custom-designed probe panel (IDT Technologies) and xGen Hybridization and Wash Kit (IDT Technologies) following manufacturer's 'tube protocol' - Final enriched library enriched with 12 cycles on-bead PCR, repurified, eluting into 12µl EB	Lumley et al[146]
HBV HEP-TILE tiled amplicon		
Nanopore library prep (sections 4.6, 4.7, 4.9, 4.9 and 6.5)	- Native barcoding kit (SQK-NBD114.96) - Protocol adapted from nCoV-2019 LoCost v3 sequencing protocol[148], with the removal of the reverse transcriptase step and substitution of a 1:10 dilution step post-PCR with a bead cleanup, to maximise the input of PCR product into the end prep reaction.	Lumley et al[149]
Illumina library prep (sections 4.6 and 4.7)	- NEBNext Ultra II FS DNA kit (NEB #E7805S) - NEBNext Multiplex Oligos for Illumina (Unique Dual Index UMI Adaptors DNA Set 1) (NEB #E7395) - As per manufacturer's instructions for fragmentation of amplicons and library prep	NEB Ultra II FS protcol[147]

CL - completion ligation; RCA - rolling circle amplification; FEN1 - flap endonuclease; gPCR - full length genomic PCR

Quality control of PCR products and sequencing libraries

I used the Qubit BR dsNDA kit (Q32853) for quantifying RCA products. I used the Qubit HS dsDNA kit (Q32854) to determine the DNA concentration of HBV PCR products and to quantify final sequencing libraries. I verified the presence of PCR product using gel electrophoresis with a 1.2% agarose gel, SYBR Safe DNA Gel Stain (S33102, Invitrogen), Hyperladder 1kb (BIO-33053) and 6X Purple loading dye (B70245 NEB). I ran the gel at 100V for 30 minutes. I used the Agilent TapeStation system (automated electrophoresis) genomic DNA ScreenTape Assay (5067-5366 and 5067-5365, sizing range is 200 to >60,000bp) to determine the length and concentration of RCA products. OGC used the High Sensitivity D1000 DNA screentape (5067-5585, 5067-5584) to QC libraries pre/post capture. I used the Agilent Bioanalyser High Sensitivity DNA kit (5067-4626 and 5067-4627, sizing range 50 - 7000bp), to determine the fragment distribution of capture and amplicon libraries.

Sequencing platforms

I sequenced Illumina libraries enriched by capture using Novaseq X 150 paired ends (for libraries prepared by OGC) or NovaSeq6000 150 paired ends (library prepared by me). I sequenced Illumina libraries enriched by HEP-TILE on a partial lane of Novaseq X 150 paired ends. I sequenced Nanopore libraries using R10.4.1 flow cells, I used the GridION platform in Kenya cohort and the MinION platform for libraries from all other settings.

2.2.3 HBV bioinformatic pipelines and data analysis

Bioinformatic pipelines

For Illumina data, Haiting Chai (Ansari group) generated consensus sequences for samples enriched by capture and HEP-TILE using a bioinformatic pipeline previously described[146]. A

base was called at the consensus level if allele depth was above x5, with HBV genome coverage calculated by determining the percentage of bases called at the consensus level.

For nanopore data, Cedric Tan (Crick/UCL PhD fellow) processed raw read data for samples enriched by RCA, calculating the number of HBV concatemers per read using the blastn v2.15.0+ alignment tool to align all Nanopore long reads generated to the to the HBV reference (X02763) and calculating the number of alignments per read that covered at least 90% of the reference length.

For analysis of samples enriched by HEP-TILE and sequenced on Nanopore, Chris Kent (University of Birmingham) developed a novel pipeline hbv-fieldbioinformatics[150] to produce consensus genomes, described in[146]. Reads were basecalled and demultiplexed with dual barcoding using dorado (0.7). Reads were then mapped against representative genomes from all genotypes[53], with the genome with the most mapped reads being selected as the primary reference. To handle the circular genome, an intermediate circular primary reference genome was created, where the sequence of the circular amplicon is appended to the 3' end of the reference. Reads were then remapped to the circular reference with minimap2 (2.26), primers were trimmed, followed by variant calling by LongShot and Medaka (0.4.5, 1.11.3). Variants and read depth for the appended region are then mapped back to their corresponding position on the original reference. A base was called at the consensus level if allele depth was over x20, with HBV genome coverage calculated by determining the percentage of called consensus bases.

Analysis of consensus sequences

I aligned consensus sequences using Mega11[151], visualised using Geneious[152] and Snapgene[153] and used IGV: Integrative Genomics Viewer[154] to view the bam files. I generated maximum likelihood phylogenetic trees using the IQTree web server[155] ultrafast bootstrap analysis with 1000 bootstrap alignments. I visualised trees using Figtree[156] with SNP-scaled branch lengths. I rooted phylogenies on the genotype H reference strain and included relevant HBV genotype and subgenotype reference sequences described by McNaughton et al. 2019[53,157].

For sequences with >50% genome coverage, I determined the HBV genotype, including recombinants, using my maximum likelihood phylogenetic trees and the online NCBI genotyping tool[158]. I called a recombinant if the recombination breakpoints occurred within a read/amplicon, to avoid mis-interpretation of mixed genotype infection as a recombinant strain. For samples with full RT coverage, I determined the presence of RAMs and VEMs using geno2pheno[159] for established RAMs and my own R code using the Biostrings[160] and Tidyverse[161] packages for putative RAMs (**Table 2.7**). I visualised amino acid proportions at positions of interest using WebLogo, Berkeley [162].

Table 2.7. RAMs which individually or in combination have been associated with HBV drug resistance.

Established RAMs (from geno2pheno[73,159])	Putative tenofovir RAMs[87]
L180M + M204VI	rtS106C + rtH126Y + rtD134E + rtM204I/V + rtL269I [163]
A181T/V	rtA194T + rtL180M + rtM204V [164]
N236T	rtR153W + rtI163V + rtL180M + rtM204V[165]
L180M + M204V/1 +/- I169T +/- V173L +/- M250V	
L180M + M204V/I +/- T184G +/- S202I/G	

RAM - Resistance associated mutation

In order to establish whether any viral polymorphisms were enriched in sequences from individuals with VBT on TFV compared to NA-naive individuals with the same subgenotype (**Chapter 6**), I compared the proportion of each amino acid between the treatment-experienced group and a comparator group of treatment-naive individuals from the same cohort and with the same genotype, and performed a Fisher's exact test with a Bonferroni correction to correct for multiple comparisons. I matched the HBV subgenotype to avoid spuriously identifying differences arising from the different distribution of genotypes within cohorts. I used the same approach to establish whether a set of polymorphisms were enriched in sequences from individuals with/without fibrosis/cirrhosis (TE >7kPa). I performed the analysis in R using the Biostrings[160] and Tidyverse[161] packages.

Chapter 3: SARS-CoV-2 WGS development and application to infection prevention control

3.1. Availability of content

This chapter was peer-reviewed and published in the Journal of Infection in 2021[166]. I am the first author of the paper, performed the SARS-CoV-2 sequencing, epidemiological vs. genomic analysis and wrote the draft of the manuscript. It has since been cited 59 times and been referenced in two policy documents[167]. In this thesis chapter, I have added additional methodological details detailing the work I performed to establish the in-house SARS-CoV-2 sequencing protocol at Oxford University Hospitals.

Study sequence data have been deposited in the European Nucleotide Archive under study accession number PRJEB43319. The multiple sequence alignment and corresponding maximum likelihood phylogeny is available from[168,169]. An interactive Microreact phylogeny is available at[170].

The epidemiological datasets analysed are not publicly available as they contain personal data but are available from the Infections in Oxfordshire Research Database[171] subject to an application and research proposal meeting the ethical and governance requirements of the Database.

3.2. Introduction

Despite robust efforts, both patients and staff acquire SARS-CoV-2 infection in hospitals. Limiting acquisition of SARS-CoV-2 by patients and staff is an IPC priority. In December 2020, at the time of this work, it was estimated that 10-40% of hospital-diagnosed COVID-19 cases were acquired in hospital[172–175]. Higher rates of seroconversion were reported in healthcare workers compared to the general population[19,176].

Distinguishing which patients have acquired infection in hospital allows potential transmission events to be investigated, and provides an evidence-base for enhancement of IPC measures to reduce transmission in the future. Epidemiological rules are frequently used for nosocomial classification and outbreak investigation for different pathogens, using spatial and temporal patient data to make assumptions about acquisition and transmission. At the time of writing this chapter, SARS-CoV-2 WGS was a recent adjunct to assist hospital outbreak investigation. Individuals infected with identical or near-identical (≤1 SNP) viruses, are more likely to be linked in a transmission chain than those with more distantly related viruses, as demonstrated by previous retrospective studies that have utilised WGS to identify nosocomial infections and outbreaks of other pathogens[177–182].

3.3. Aims

Aim 3.1. To investigate whether SARS-CoV-2 WGS can enhance epidemiological investigation and classification of healthcare-associated SARS-CoV-2 acquisition by confirming/excluding nosocomial acquisition in patients

Aim 3.2. To investigate the role of WGS in understanding SARS-CoV-2 transmission patterns within hospitals

3.4. Methods

3.4.1 Study design, setting and participants

The study population consisted of all inpatients (adults and children) and staff who tested positive for SARS-CoV-2 at OUH, Oxford, UK between 17-November-2020 to 5-January-2021, a period of exponential spread with the emergence of the first 'alpha' VOC of SARS-CoV-2 (originally called B.1.1.7)[183,184] (**Figure 3.1**). SARS-CoV-2 PCR was performed on combined nasal and oropharangeal swabs (see **Chapter 2** for more details on study setting, testing strategy and PCR assays). Ward admission and discharge dates, including ward movements, were available for all patients from 14 days before the first positive PCR, and the work location for those staff working exclusively or predominantly on a single ward.

National public health guidance for COVID-19 IPC was followed throughout the study, including the use of patient pathways, PPE, symptomatic and asymptomatic staff and patient testing, summarised below[185]. At the time, recommendations were issued by Public Health England (PHE); this has subsequently become the UK Health Security Agency (UKHSA). The study spanned the roll out of the COVID-19 vaccination programme in the UK, which began on 8th December 2020.

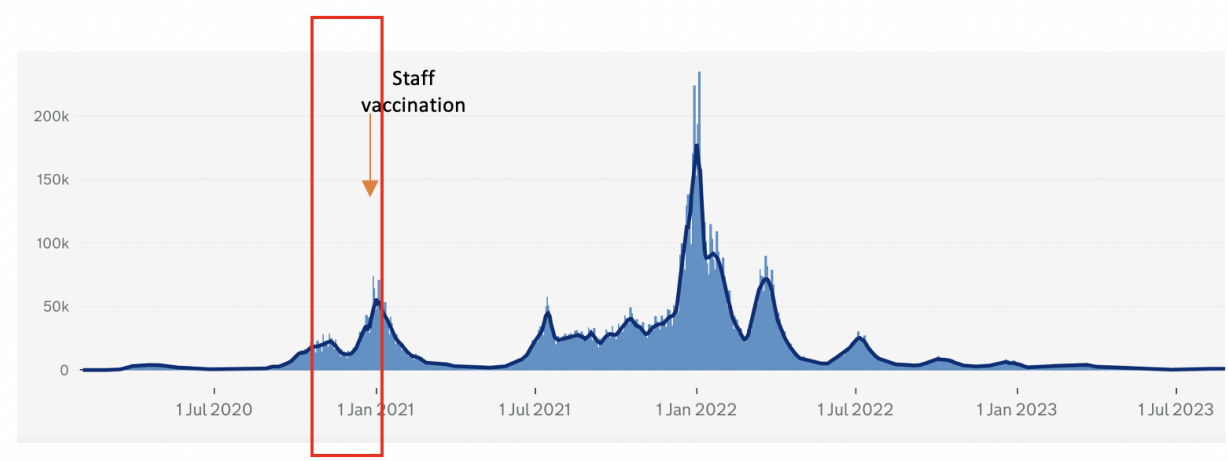


Figure 3.1. Epidemic curve showing the daily count of UK confirmed SARS-CoV-2 cases by specimen date in relation to the study period (source gov.uk). Red box indicates the study period covered by this chapter.

3.4.2 SARS-CoV-2 WGS

During this study, I established an 'in-house' SARS-CoV-2 sequencing capability in the main hospital adjacent to the microbiology lab and was able to batch and sequence samples generating genomes within 4-5 working days of a positive PCR result. I sequenced residual extract from the clinical microbiology laboratory.

Following a literature review of existing SARS-CoV-2 WGS protocols, I selected the ARTIC LoCost protocol[114], a widely used amplicon-based enrichment method followed by Nanopore sequencing. I compared two sets of primers (the "ARTIC v3" primers which produce short 400bp amplicons[186] and "Midnight" primers[187] which produce longer 1200bp amplicons), on a set of 10 known positive clinical samples in duplicate at a range of cycle threshold (Ct) values. The shorter ARTIC v3 amplicons performed consistently well across Ct values. The

Midnight primers performed poorly at higher Ct values due to the disproportionate impact of single long amplicon dropout on genome coverage, therefore I took the shorter ARTIC v3 primers forwards for further sequencing work. Due to persistent challenges with sequencing samples at low viral load, I further improved genome coverage by reducing the amount of the final library discarded, capitalising on the possibility of washing and re-loading a flow cell part way through a sequencing run. Details of the final sequencing protocol are included in **Chapter 2**.

3.4.3 Epidemiologic and genetic classification of community v.s. hospital acquisition

I took a four stage approach to classifying whether inpatients had hospital-acquired or community-acquired SARS-CoV-2 infection:

i) Standard epidemiological classification

Initially I classified all inpatient infections according to NHS England and NHS Improvement definitions[188] below, which are based on the number of days between admission and the first positive PCR:

- **Community-Onset**, PCR-positive ≤2 days after hospital admission/attendance
- **Hospital-Onset Indeterminate Healthcare-Associated**, PCR-positive 3-7 days after hospital admission (hereafter referred to as “Indeterminate nosocomial”)
- **Hospital-Onset Probable Healthcare-Associated**, PCR-positive 8-14 days after admission (“Probable nosocomial”)
- **Hospital-Onset Definite Healthcare-Associated**, PCR-positive ≥15 days after admission (“Definite nosocomial”)

ii) Enhanced epidemiological classification

I then re-classified inpatients, taking into account prior negative PCR results (available as a result of admission screening, weekly ward screening and symptomatic testing) and results from admissions in the 14 days prior to diagnosis to determine whether additional support existed for nosocomial acquisition.

iii) Epidemiological evidence of transmission

Next, I tested if there was epidemiological evidence of a potential source case for each new definite/probable/indeterminate patient and staff infection. I denoted individuals acquiring SARS-CoV-2 as “recipients", and those transmitting infection as “donors”. In this context, a “plausible donor” for a recipient, is identified by the donor and recipient being present on the same ward (“ward contact”), during the donor’s infectious period and the recipient’s incubation period. I defined “hospital contact” as the presence in the same hospital on the same calendar day, during the donor’s infectious period and the recipient’s incubation period.

For the purpose of identifying plausible transmission events, I defined indicative incubation periods as 1-14 days prior to a positive PCR test[15]. I defined infectious periods from 4 days before to 7 days after a positive PCR for patients, and 4 days before to the day of the positive PCR test for staff (reflecting that staff isolated at home for 10 days following a positive test).[189] Mean serial intervals, i.e. the duration between the symptom-onset time of in a transmission donor and recipient, have been estimated at 4-7 days, here I use 5 days[190,191].

iv) Combined epidemiological and genomic evidence of transmission

I then evaluated how many of these epidemiologically linked cases had virus consensus sequences within ≤1 SNPs of each other, i.e. had genomic evidence to support transmission. The estimated evolutionary rate for SARS-CoV-2 is 1.1×10^{-3} substitutions/site/year, which

equates to 2-3 substitutions per genome per month[192]. This evolutionary rate is slow compared to the rapid transmission of cases in the community and in hospital, with a serial interval of approximately 5 days[190]. This leads to the possibility that two people may be infected with identical viruses by chance, rather than due to direct person-to-person transmission. Using a cumulative Poisson distribution, with approximations of one mutation every 2 weeks and a serial interval of 5 days, there is 70% chance of no new SNPs per transmission, 95% of 0-1 SNPs, and 99% of 0-2 SNPs, therefore the higher the SNP threshold set for considering transmission, the more true transmissions are captured. However, the extent of population-level diversity also needs to be considered, i.e. the probability of two randomly chosen sequences being within 0, 0-1, or 0-2 SNPs within this data set was 0.9%, 2.1% and 3.5% respectively. Therefore a plausible genetic relationship was defined as ≤1 SNP between two samples in this study, an association close enough to support transmission whilst minimising over-calling of linkage in non-nosocomial cases.

3.4.4 Outbreak analysis

i) Epidemiological outbreaks

I searched for epidemiologically defined outbreaks involving infected patients and staff. I defined epidemiological outbreaks following PHE guidance:[193] ≥2 cases of COVID-19 in patients or hospital staff 'associated with the same setting', with ≥1 case (if a patient) meeting the definition of probable/definite nosocomial infection, ending when no cases were diagnosed for 28 days. Here I defined 'associated with the same setting' as a ward contact. I only included community-onset cases as part of an outbreak if they could have plausibly seeded the outbreak (i.e. their diagnosis preceded the first staff or nosocomial patient case on that ward), and not if admitted during an ongoing outbreak.

ii) Genomic outbreaks

I defined “genomic outbreaks” as for epidemiological outbreaks with the additional requirement for individual viral consensus sequences to be genomically linked (sharing ≤1 SNPs). I defined “genomic clusters” as for genomic outbreaks, but without the requirement for ≥1 definite/probable nosocomial case.

3.5. Results

3.5.1 Overview of study population

From 17-November-2020 to 5-January-2021, 1132 individuals (803 inpatients, with 1104 admissions, and 329 staff) were newly diagnosed with PCR-confirmed SARS-CoV-2 infection at OUH (**Figure 3.2**). The median inpatient age at diagnosis was 67 years (IQR 49-81, range 0-102), 43% were female. 188/803 (23%) inpatient infections were classified as nosocomial (definite, probable or indeterminate). Length of stay after the first positive PCR was a median of 6 days in community-onset cases vs. 8 days in nosocomial cases (Kruskal-Wallis $p<0.001$). All-cause mortality in patients within 28 days of a first positive SARS-CoV-2 PCR was 14% after community-onset and 23% after nosocomial infection ($p<0.001$) (**Table 3.1**). Swabs from 764/1132 (67%) SARS-CoV-2 PCR-positive individuals were successfully sequenced, including 116/188 (62%) nosocomial cases and 261/329 (79%) staff cases (**Tables 3.2, 3.3, Figure 3.3**).

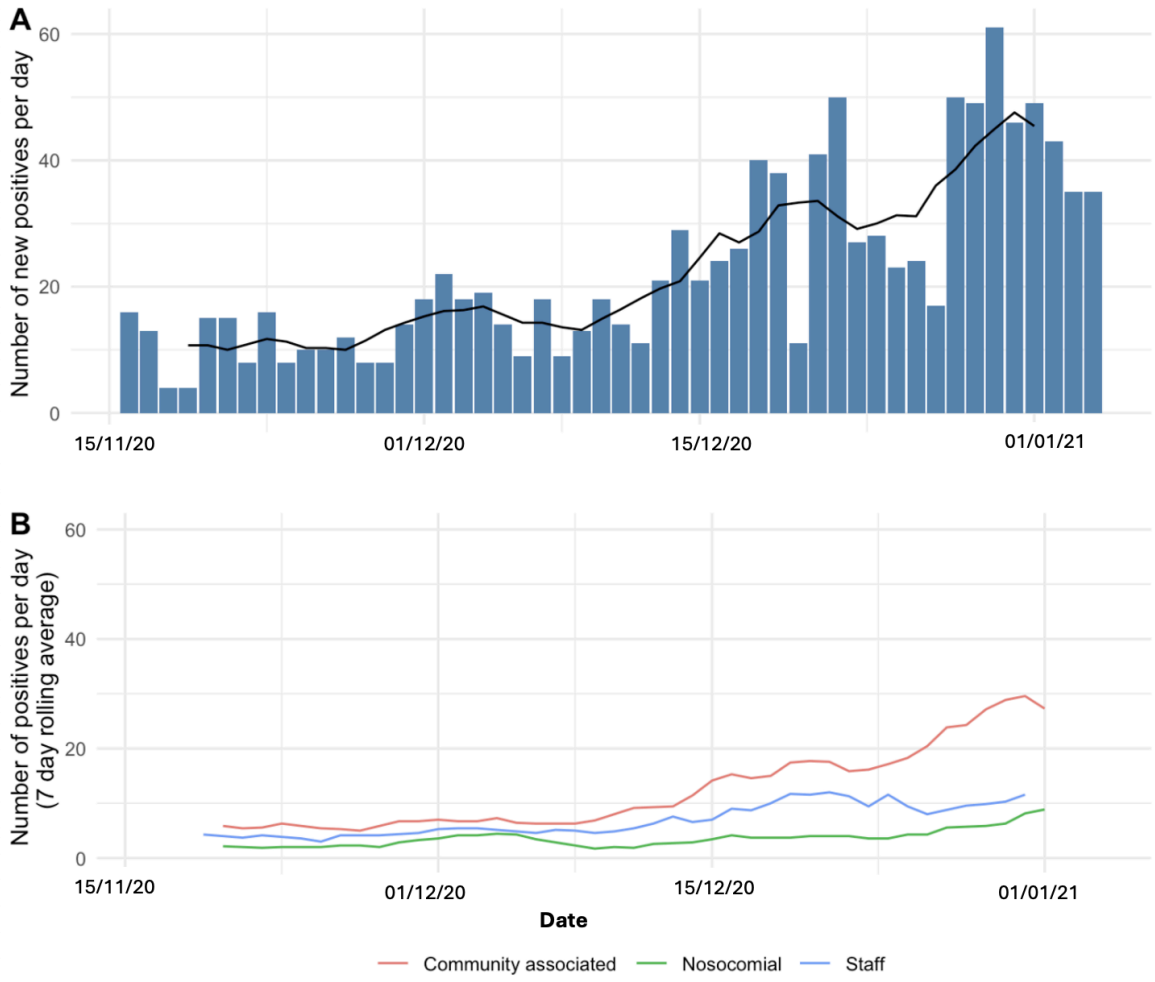


Figure 3.2. Epidemic curves showing number of new SARS-CoV-2 PCR staff and inpatient positives at Oxford University Hospitals from 17 November 2020 - 5 January 2021 based on clinical microbiology laboratory diagnosis by PCR. A) total number of staff and inpatient positives (blue bars showing new positives per day, black line shows 7 day rolling average) B) SARS-CoV-2 positives split by staff and patients (nosocomial (definite, probable and indeterminate) and community-onset according to national surveillance definitions), lines show 7 day rolling average.

Table 3.1. Characteristics of 803 inpatients testing SARS-CoV-2 PCR positive between 17 November 2020 and 5 January 2021 at Oxford University Hospitals.

	All patients	Nosocomial	Community-onset	Kruskal Wallis p-value
Total number	803	188 (23%)	615 (77%)	NA
Sex **Male (n; %)** **Female (n; %)**	 454 (57%) 349 (43%)	 109 (58%) 79 (42%)	 345 (56%) 270 (44%)	0.65
Age (median; IQR)	67 (49-81)	76 (62-86)	63 (46-80)	<0.001
PCR Ct value (median; IQR)	20 (14-27)	19 (14-28)	20 (15-27)	0.15
Length of stay before first SARS-CoV-2 positive PCR (median; IQR)	1 (1-3)	9 (5-15)	1 (1-1)	<0.001
Length of stay after first SARS-CoV-2 positive PCR (median; IQR)	6 (2-13)	8 (4-16)	6 (2-11)	<0.001
28 day mortality (n; %)	129 (16%)	43 (23%)	86 (14%)	<0.001
Successful WGS (n; %)	503/803 (63%)	116/188 (62%)	387/615 (63%)	0.44

IQR - interquartile range; PCR - polymerase chain reaction; Ct - cycle threshold, WGS - whole genome sequencing

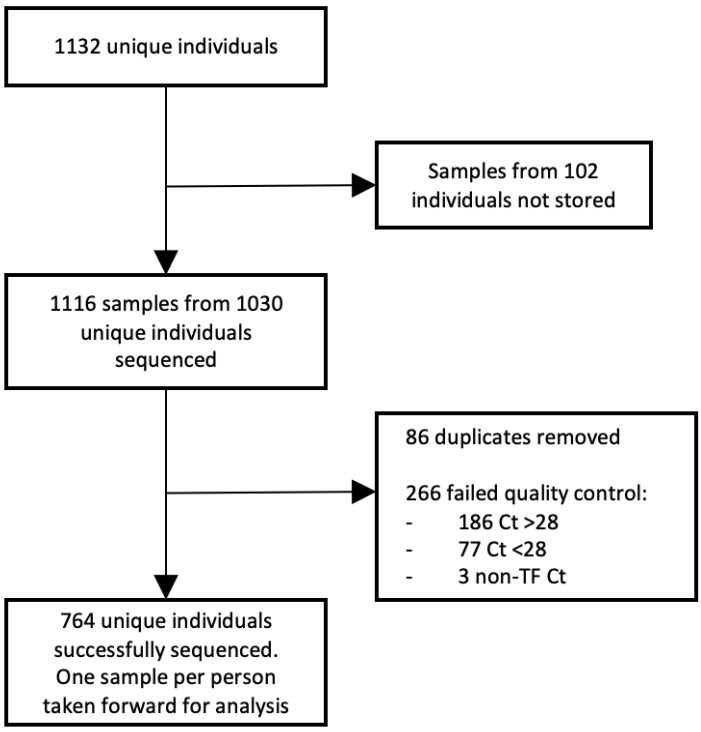


Figure 3.3. Flow diagram showing the numbers of SARS-CoV-2 samples available for sequencing and being successfully sequenced. Where duplicate samples from an individual were sequenced, the sample with highest percentage SARS-CoV-2 genome coverage was selected for analysis. TF = Thermofisher (only Ct values for the most commonly used Thermofisher PCR platform are reported).

Table 3.2. SARS-CoV-2 sequencing pass rates by the standard epidemiological nosocomial classification and Thermofisher Ct value.

Epi classification	Total	Pass			Fail			Not available		
		TF Ct <28	TF Ct >28	Non TF	TF Ct <28	TF Ct >28	Non TF	TF Ct <28	TF Ct >28	Non TF
Definite	51	24	3	0	4	16	0	2	0	2
Probable	58	33	9	0	2	11	0	1	0	2
Indeterminate	79	39	8	0	6	22	1	1	2	0
Community-onset	615	361	26	0	56	93	0	21	35	23
Staff	329	238	20	3	9	44	2	1	1	11

Epi - epidemiological; Ct - cycle threshold.

TF Ct <28 = Ct value of sequenced sample was less than 28

TF Ct > 28 = Ct value of sequenced sample was more than 28

Non TF = indicates the PCR was not performed on the Thermo Fisher platform (to allow comparison, only Thermo Fisher Ct values are reported).

Table 3.3. Demographics of inpatients testing PCR positive for SARS-CoV-2 infection at Oxford University Hospitals between 17 November 2020 and 5 January 2021, representing those for whom SARS-CoV-2 was sequenced vs non-sequenced

	All patients	Sequenced	Not sequenced
Total number	803	503 (63%)	300 (37%)
Sex **Male (n; %)** **Female (n; %)**	 454 (57%) 349 (43%)	 292 (58%) 211 (42%)	 162 (54%) 138 (46%)
Age (median; IQR)	67 (49-81)	69 (51-81)	63 (43-80)
Lowest Ct (median; IQR)	20 (14-27)	17 (14-21)	28 (22-32)
Length of stay before first SARS-CoV-2 positive PCR, days (median; IQR)	1 (1-3)	1(1-2)	1 (1-5)
Length of stay after first SARS-CoV-2 positive PCR, days (median; IQR)	6 (2-13)	7 (2-14)	6 (2-11)
28 day mortality (n; %)	129 (16%)	97 (19%)	32 (11%)

IQR - interquartile range; Ct - cycle threshold

3.5.2 Epidemiological classification

Nosocomial classification

Based on standard national definitions, I classified 188/803 (23%) inpatient infections as nosocomial, subgrouped as definite (n=51), probable (n=58) or indeterminate (n=79). In the UK, patients who acquired SARS-CoV-2 infection in hospital but were discharged before testing positive are not reported as nosocomial and so are not accounted for in these numbers, in part because community testing results are not routinely available to hospitals.

"Enhanced" nosocomial classification

Admission screening (within 24 hours) was performed in 916/1104 (83%) of admissions. All 51 definite cases had prior-negative PCRs earlier in the same admission, providing additional support for nosocomial acquisition. 56/58 (97%) probable cases had prior-negative PCRs. Although described as indeterminate, 53/79 (67%) of those diagnosed 3-7 days after admission had ≥1 prior negative sample (**Table 3.4**). Indeterminate cases with a prior-negative PCR were diagnosed later during admission (median day 5, range 3-7), than those without a prior-negative (median day 3.5, range 3-7; p=0.001). Therefore the greatest uncertainty around nosocomial acquisition exists for the 26 indeterminate cases without prior-negative PCRs and relatively short intervals between admission and first positive PCR. By definition, those without a prior-negative PCR had no prior PCR tests obtained in the same hospital admission, i.e. admission testing was not done.

Amongst the 615 community-onset cases, a retrospective look-back at the 14 days prior to SARS-CoV-2 diagnosis revealed 89/803 (11%) had prior hospital admissions during which SARS-CoV-2 infection could have been acquired, and 69/89 (78%) of these had ward contact with a plausible donor during that admission, suggesting reporting definitions based only on current admissions under-estimate the extent of nosocomial infection.

I identified similar proportions of epidemiologically plausible donors for 46/51 (90%) definite cases, 55/58 (95%) probable cases and 75/79 (95%) indeterminate cases (exact p=0.52). The proportion of cases with plausible donors identified did not differ in probable or indeterminate nosocomial cases with/without prior negative PCRs (p>0.90 and p=0.59 respectively).

Table 3.4. Epidemiological and genomic evidence for nosocomial acquisition of SARS-CoV-2 based on classification of 803 PCR-positive hospital in-patients

	Definite nosocomial		**Probable nosocomial**		**Indeterminate**		**Community-onset**	
Epidemiological evidence								
Standard epidemiological classification*	51		58		79		615	
Enhanced epidemiological classification provides support for nosocomial acquisition (n,%)**	**Yes**	**No**	**Yes**	**No**	**Yes**	**No**	**Yes**	**No**
	51 (100%)	0	56 (97%)	2 (3%)	53 (67%)	26 (33%)	89 (14%)	526 (86%)
Donor/recipient closest epidemiological link identified								
Ward contact	46	0	53	2	51	24	69	0
Hospital contact	5	0	3	0	2	2	20	0
Combined epidemiological and genomic evidence								
Recipient and one or more epidemiological donor(s) sequenced (n,%)^								
Ward contact	25/46 (54%)	0	37/53 (70%)	1/2 (50%)	29/51 (57%)	15/24 (63%)	37/69 (54%)	0
Hospital contact	1/5 (20%)	0	2/3 (67%)	0	1/2 (50%)	1/2 (50%)	5/20 (25%)	0
Genomic linkage confirmed (n,%)^^								
Ward contact	17/25 (68%)	0	32/37 (86%)	1/1 (100%)	22/29 (76%)	6/15 (40%)	17/37 (46%)	0
Hospital contact	1/1 (100%)	0	1/2 (50%)	0	1/1 (100%)	0/1 (0%)	0/5 (0%)	0

* Nosocomial classification according to national Public Health England definition

**Enhanced epidemiological classification for nosocomial acquisition, using prior negative PCR in the same admission, or previous hospital admission during the incubation period to provide additional support for nosocomial acquisition. Figures demonstrate number and % of nosocomial category.

^ Figures presented as number sequenced over total number of recipients with one or more ward or hospital contacts identified

^^ Figures presented as number genomically linked over total number of recipients sequenced, for ward contacts or hospital contact

Epidemiological outbreak identification

Applying an epidemiological outbreak definition considering all ward overlaps led to the identification of three within-hospital outbreaks, the largest containing over 700 individuals, highlighting that existing definitions were unworkable when inpatient prevalence was high. Therefore, to more closely replicate IPC practice and provide more interpretable data, I restricted the definition of an epidemiological outbreak to only include ward overlaps with patients and staff on the ward of nosocomial diagnosis. A total of 246 individual infections (46 definite, 53 probable, 56 indeterminate, 7 community-onset and 84 staff), occurred in one of 25 outbreaks on 24 different wards (**Figure 3.4**). The median outbreak size was 8 (IQR 3-12, range 2-32); 99/109 (91%) definite and probable cases occurred in outbreaks.

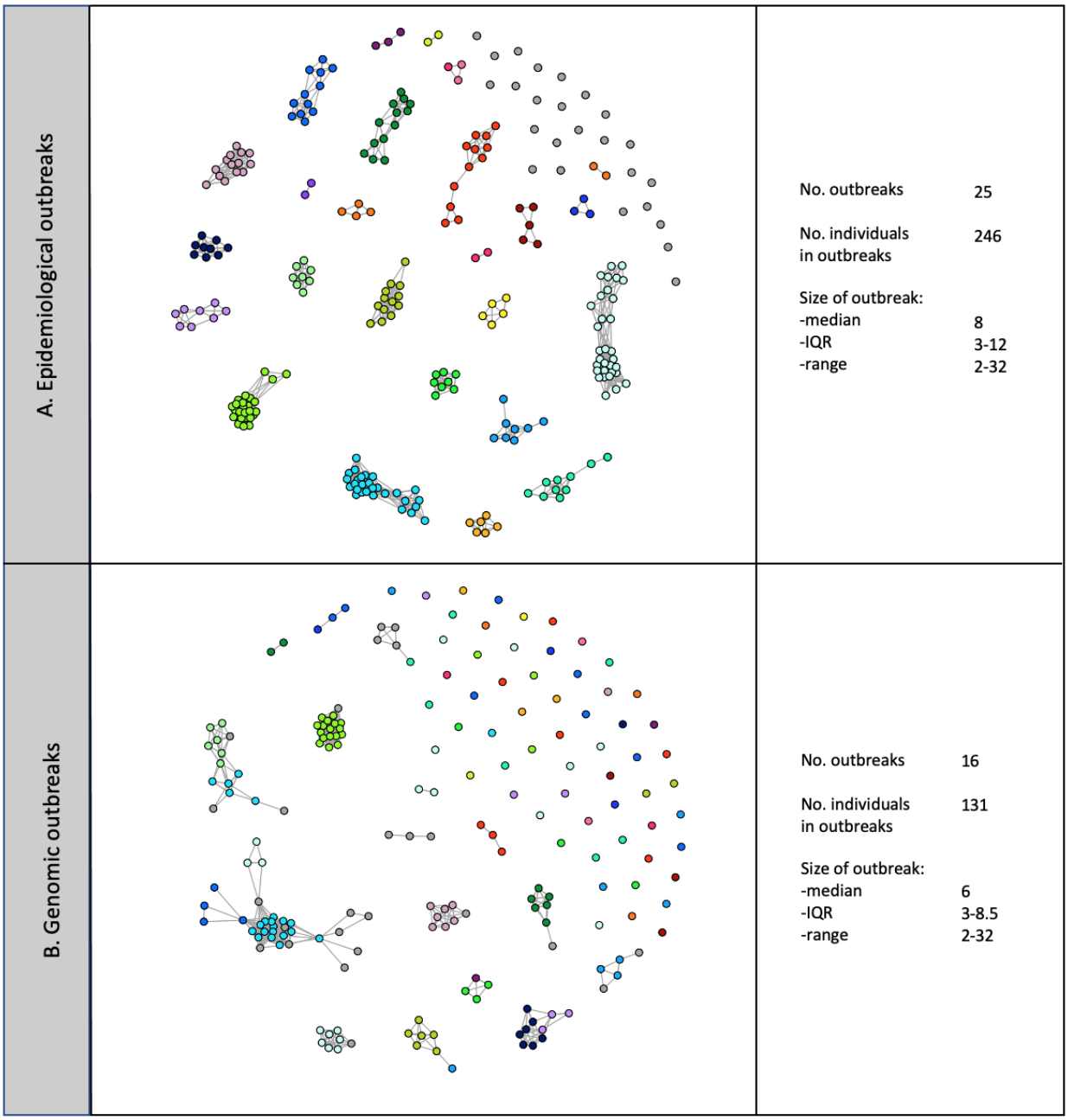


Figure 3.4. SARS-CoV-2 outbreaks at Oxford University Hospitals between 17 November 2020 and 5 January 2021 containing at least one definite or probable nosocomial case.

A) Using epidemiological data alone (nodes are linked purely using ward-based contacts) isolated grey nodes indicate individuals in a genomic but not epidemiologically-defined outbreak,

B) Using both epidemiological data and genomic data (nodes are linked both epidemiologically and genomically), isolated nodes indicate individuals in an epidemiological but not genomic outbreak.

Each node represents an individual, all individuals in an epidemiological or genomic outbreak are shown in panel A with the sequenced subset in panel B. Node colours indicate the epidemiological group, grey nodes were not assigned to an epidemiological group. Lines indicate ward contact within an outbreak, line length is insignificant. This demonstrates that epidemiological outbreaks consist of multiple genomic outbreaks and individual introductions, and conversely genomic outbreaks span multiple wards/epidemiological outbreaks. 69/176 (39%) nosocomial cases were not sequenced.

3.5.3 Can genomics help to confirm/exclude nosocomial acquisition?

Genomics provides confirmatory evidence of nosocomial acquisition for most cases classified as nosocomial

Of the epidemiologically identified nosocomial acquisitions with plausible ward donors, I successfully sequenced 107/176 (61%) samples alongside ≥1 plausible donor samples. I found genomic support for nosocomial acquisition for 78/107 (73%) sequenced infections (17/25 (68%) definite, 33/38 (87%) probable and 28/44 (64%) indeterminate). Three further individuals with nosocomial infection, but without ward contacts, were genomically linked to a hospital-level contact (**Table 3.4**), i.e. plausibly representing transmission via an undiagnosed/unsequenced patient, staff contact in communal areas, staff cross-covering multiple wards or through environmental contamination.

Genomics improves precision for cases without prior-negative PCR results

Genomics helped clarify uncertainty around cases without a prior-negative PCR (**Table 3.4**). In the probable group, two individuals lacked a prior-negative PCR; one was sequenced alongside a donor and confirmed as genomically-linked (≤1 SNP), and therefore likely nosocomially-acquired. In contrast, in the indeterminate group, 26 individuals lacked a prior-negative PCR. 15/24 (63%) were sequenced alongside ≥1 potential donors; only 6/15 (40%) were genomically-linked. Hence absence of a prior-negative PCR in the indeterminate

group was associated with a lower likelihood of nosocomial-acquisition, but sequencing did support some of these infections having a nosocomial source.

Genomics can confirm nosocomial acquisition in "community-onset" cases

Of the 69 individuals with community-onset infection with a prior hospital admission and a plausible donor, 37 were sequenced alongside ≥1 plausible donor(s). 17/37 were genomically-linked, indicating 17 additional infections previously categorised as "community-associated" were plausibly nosocomially acquired (**Table 3.4**).

Undiagnosed/unsequenced individuals limit utility of genomic data

Amongst the 116 nosocomial cases sequenced, 13 (11%) were genetically-linked to ≥1 other case within 0-1 SNPs but with no documented ward or hospital contact (either patient or staff). These may represent community acquisition in the case of indeterminate cases, but may also be due to undiagnosed/unsequenced individuals providing the missing epidemiological link, e.g. due to incomplete admission and ward-based patient screening and undiagnosed staff cases.

Genomic data was unable to provide confirmation of nosocomial acquisition for 22/116 (19%) of sequenced nosocomial cases. Although I can conclude that these cases were not linked to any of the other cases sequenced, I cannot use this information to exclude nosocomial acquisition from undiagnosed/unsequenced individuals, due to incomplete sampling/ sequencing.

3.5.4 Can genomics improve understanding of transmission patterns in hospitals?

Epidemiological outbreaks often contain multiple genomically-distinct introductions

I used genomic data to refine the epidemiologically-defined outbreaks (**Figure 3.4**). Of the 246 staff and patients in an epidemiological outbreak, I sequenced 171 (26/46 definite, 37/53 probable, 33/56 indeterminate patients, 2/7 community-onset and 73/84 staff); 21 of the 25 epidemiological outbreaks had ≥2 members sequenced and are described further. One epidemiological outbreak was confirmed genetically as a single outbreak (each case within 0-1 SNPs of another) and in 7 'outbreaks' no individuals were genetically linked (all ≥2 SNPs from all other cases). In the remaining 13 epidemiological outbreaks there was a mix of genetically-linked and unlinked cases; 9 consisted of a single genetically-distinct outbreak in addition to unlinked cases, 3 had two distinct genomic outbreaks and one ward had evidence of 3 genomic outbreaks. Overall 116/171 sequenced cases from epidemiologically defined outbreaks were confirmed to be in a genomically-supported outbreak (17/26 (65%) definite, 34/37 (92%) probable, 26/33 (79%) indeterminate, 1/2 (50%) community-onset and 38/73 (52%) staff). This highlights that epidemiological investigation may overestimate the size of outbreaks, which often occur alongside genetically-distinct introductions.

Most cases do not lead to onwards hospital transmission

Considering the population studied as a whole, rather than just those in an epidemiological outbreak as above, of the 764 individuals with samples sequenced, 200 were placed in one of 43 genomic clusters on the basis of being linked to at least one other case within 0-1 SNPs and 564 were singletons. Therefore during the period of study, SARS-CoV-2 was introduced to OUH on at least 607 occasions, with evidence of onward transmission in 43 clusters (7% of introductions) (**Figure 3.5**). The median cluster size was 2 (range 2-32). Of

the 43 clusters, 17 contained both staff and patients, 16 patients only and 10 staff only. 16/43 genomic clusters were classified as genomic outbreaks (i.e. contained ≥1 definite/probable nosocomial case). Compared to the epidemiological estimate that 91% of nosocomially-acquired cases were linked to outbreaks, combining epidemiological and genomic data suggested 52/69 (75%) of all sequenced definite/probable nosocomial cases occurred in one of 16 genomic outbreaks, whereas 26% occurred as genomic singletons, ≥2 SNPs from any other case.

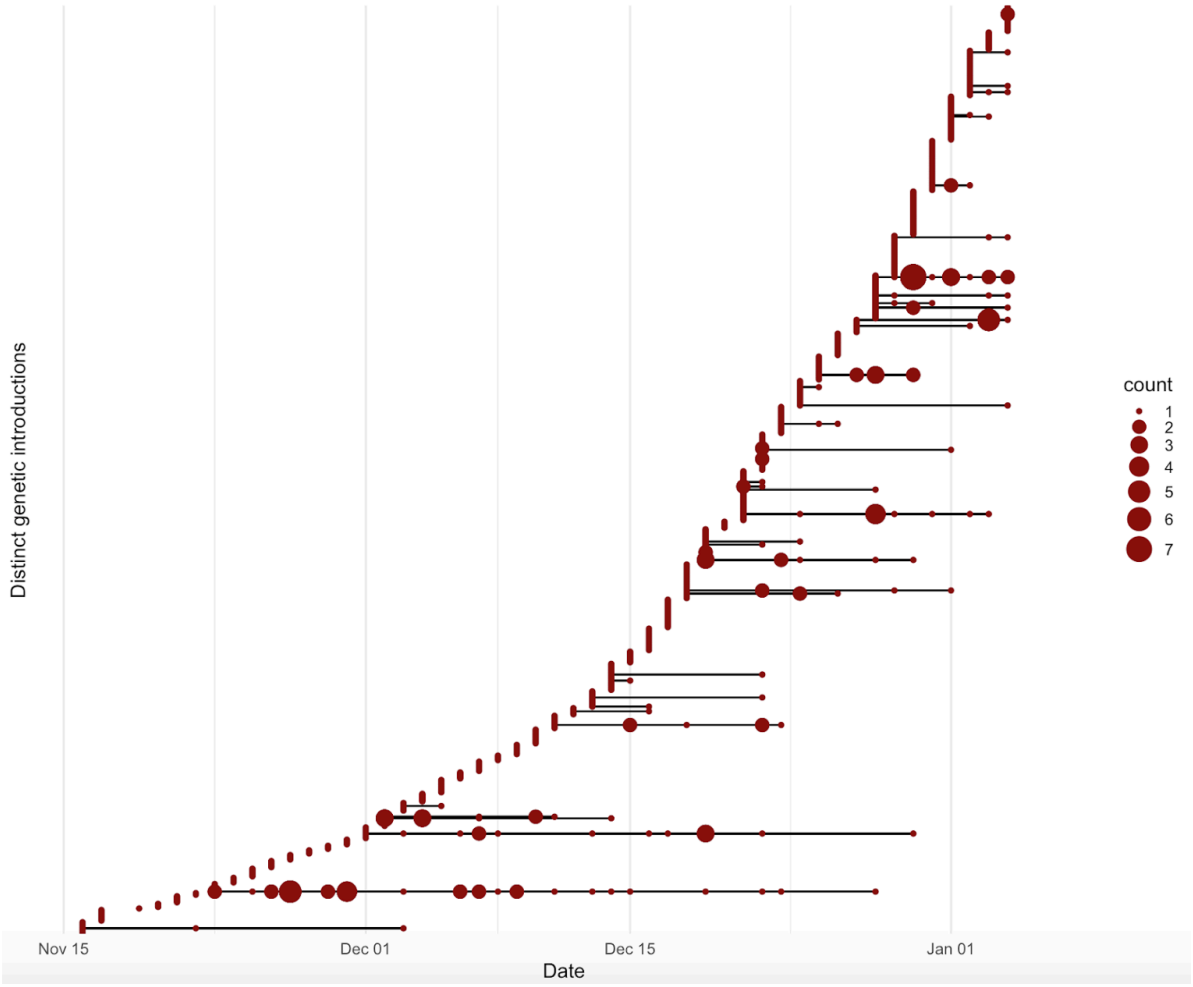


Figure 3.5. Timing and size of SARS-CoV-2 genetic clusters at Oxford University Hospitals between 17 November 2020 and 5 January 2021. SARS-CoV-2 was introduced to OUH on at least 607 occasions, with evidence of onward transmission on 43 occasions. Isolates >1 SNP from any previous sample were defined as distinct introductions and are plotted on separate horizontal lines, according to the date of the first positive sample in OUH for each individual. The size of the red dot indicates the number of individuals diagnosed on each day.

Use of hospital-level ward data, accounting for all patient moves before and after testing PCR positive, led to identification of an unfeasibly large epidemiological outbreak of over 700 individuals. However, using these data in combination with WGS provides a more plausible identification of 15 additional individuals linked to outbreaks, who were missed by ward-based application of the outbreak definition due to patient ward moves during their incubation period, highlighting that outbreaks can span multiple wards.

Only 7/25 epidemiologically-defined outbreaks started with a known community-onset case, I successfully sequenced 2/7, and confirmed only 1 to be genetically-related to subsequent cases within 0-1 SNPs. Despite the partial sequencing of community-onset cases, these data are consistent with limited direct patient-patient spread from known community-onset SARS-CoV-2 infected patients. Approximately two-thirds of staff infections were genetically distinct in this dataset, with 170/261 (65%) >1 SNP different to all other cases, across 90 work locations. Although these cases occurred on wards with existing outbreaks, they were more common in areas with transient patient contact e.g. outpatient areas and dialysis units.

The distribution of genomic clusters differed by hospital site; consistent with the extent of exposure to COVID-19 admissions. Only isolated/single cases occurred at hospital “D” and only two clusters in hospital “C” (one staff pair and one trio containing 1 staff member and 2 patients). In contrast, I observed multiple larger clusters in hospitals “A” and “B” (notably, the proportion of cases sequenced was the same across all sites). In addition to differences in COVID-19 case load/infectious pressure, other factors may have played a role, such as: patient pathways including co-location of non-COVID-19 and COVID-19 cohort wards, estates/facilities, including number of patient side rooms and ventilation, differences in staff mobility between COVID-19 and non-COVID-19 wards, staff facilities (communal/break areas) and adherence to social distancing.

Two patterns of nosocomial acquisition seen

Broadly I observed two patterns of nosocomial acquisition; patterns are shown on a representative example phylogeny in **Figure 3.6**.

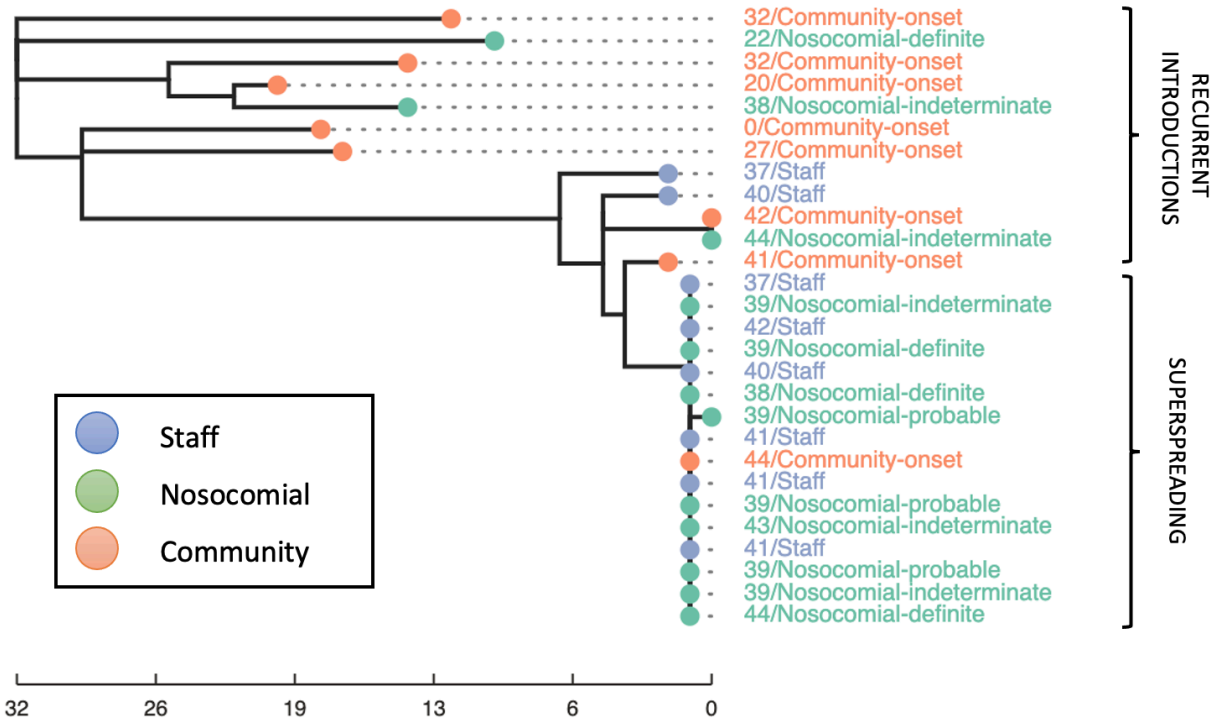


Figure 3.6. Example phylogeny demonstrating a superspreading event and recurrent introductions on a single ward at OUH hospital "B". The node and label colour indicates broad epidemiological classification (community-onset, nosocomial, staff). The tip label gives the day of the outbreak the individual tested positive followed by the full epidemiological classification. The scale bar represents SNP distance. All cases were classified as part of the same epidemiological outbreak, however WGS reveals multiple introductions. The "community-onset" case diagnosed on day 44 of the outbreak, had a previous hospital admission with exposure on this ward during a superspreading event.

i) Superspreading events

These are characterised by a rapid accumulation of multiple cases within 1-2 serial intervals, e.g. >5 cases within a 7-10 day period, implying multiple transmissions per serial interval. 8 such events occurred during this study, 7 at hospital “A”, and one at hospital “B”, typically occurring on non-COVID-19 wards, with open bays, involving both staff and patients, and in specialties with patients highly dependent on nursing care (e.g. trauma, acute medicine, neurology). The median outbreak size was 9 (range 7-32) and median duration 12 days (range 7-35 days). Although infrequent, these 8 superspreading events accounted for 80% of cases linked to a genomic outbreak.

With a serial interval of 5 days, some outbreaks may represent exposure to a single superspreading infection, but most are subsequently propagated amongst staff/patients. Incomplete sampling and asymptomatic individuals without symptom onset dates prevents confident identification of the source of each outbreak, however, in two clusters, staff cases preceded patient cases, so staff could have acted as an index events. In the remaining six outbreaks, there were no cases diagnosed prior to the first definite/probable nosocomial case to act as a plausible index, therefore the outbreak was likely seeded by an undiagnosed or unsequenced patient/staff/visitor. No outbreaks were seeded by direct patient-patient transmission from known positive patients, however I cannot exclude a non-sequenced cross-covering staff member providing the missing epidemiological link, by acquiring infection from a known positive patient and seeding an outbreak on a non-COVID ward.

ii) Recurrent introductions

These are characterised by slow “rumbling” accumulation of nosocomial and staff cases on a ward, on both non-COVID and mixed wards with side rooms accommodating COVID and non-COVID patients (**Figure 3.6**). They consist of multiple introductions of distinct viral

variants over a more prolonged period of time, giving the appearance of a slowly progressing outbreak, but with no, or minimal, onward transmission within the unit. Genomic data is required to distinguish recurrent introductions from genomic outbreaks.

Recurrent introductions involving one or more definite/probable case occurred on 6 different wards across hospitals “A”, “B” and “D”. Each mimicked an outbreak with between 3-6 staff and nosocomial cases occurring on the ward over 2-6 week periods, however all were genetically distinct introductions.

3.6. Discussion

In this retrospective cohort study of healthcare-associated SARS-CoV-2 transmission using combined epidemiological and sequencing data, I present key findings that challenge the current surveillance definitions and reveal most nosocomial transmission occurs from a relatively limited number of highly infectious individuals.

My findings suggest that the majority of cases of SARS-CoV-2 infection that occurred after >7 days in hospital were nosocomially acquired, for example 107/109 probable/definite cases had ≥1 prior-negative PCR test in the same admission, and most had plausible ward-based sources for their infection. However, surveillance definitions (based on prior hospital stays of >7 days) likely under-estimated the extent of acquisition in hospital. In the UK, nationally reported nosocomial figures exclude indeterminate cases (diagnosed on day 3-7 of their hospital stay); however several of these cases in our study had plausible exposure to infectious patients, and genomic-linkage with other cases, supporting acquisition in hospital, particularly for cases diagnosed on days 5-7. Furthermore, surveillance definitions

considering only the current hospital stay, as in the UK, do not capture nosocomial acquisition during a recent prior hospital stay, for which I found both epidemiological and genomic evidence. Consideration should be given to revising surveillance definitions to account for prior-negative tests and infections diagnosed <7 days into admission, although on review at the time of writing my thesis in 2024, no changes to definitions had been made.

Consistent with defining most SARS-CoV-2 infections within two days of admission as community-acquired, genomics demonstrated most cases in staff and patients were genomically-distinct from all others in the hospital; there were 607 genomic clusters within the 764 samples sequenced. This is similar to WGS-based findings in other healthcare-associated infections over the last decade[194,195]. However, in contrast to other nosocomial infections, I found evidence that most nosocomial acquisition occurs in explosive superspreading events, with clusters of genomically-related cases occurring in short time periods, as observed by others for SARS-CoV-2 in both community and hospital settings[179,196–199].

WGS added most value when investigating outbreaks during periods of high SARS-CoV-2 prevalence, given high rates of ward-based contact with infected patients. The majority of epidemiologically-defined outbreaks consisted of multiple genomic introductions with some smaller genomic clusters. The role of staff in outbreaks is overestimated from epidemiological data alone, with genomics confirming only 52% of staff epidemiologically placed in an outbreak were genomically-linked, and the majority of sequenced staff cases were genomic singletons (≥2 SNPs from any other case). Additionally, in hospitals not routinely admitting COVID-19 patients, rates of transmission were low, suggesting that isolated acquisition from staff is relatively uncommon, and that transmission requires a ‘perfect storm’ of mixed COVID

and non-COVID wards, emergency admissions and dependent patients accommodated in bays.

The main limitations of genomic data were two-fold. Firstly, although epidemiological data were available for all patients, genomic data were limited by sample availability and difficulty of generating sequences from samples with low viral loads (high Ct values). Here, I successfully sequenced 67% of relevant samples, in line with other similar hospital cohorts (20-70%)[177,181,200]. Due to the gaps in the dataset, genomic data does not enable nosocomial acquisition to be ruled out. Incomplete hospital sequencing datasets suffer from an 'absence of evidence' when attempting to exclude nosocomial acquisition, which should not be mistaken for 'evidence of absence' of nosocomial acquisition. This may be mitigated by integrated community epidemiological and genomic datasets, or by further optimising sequencing yields.

Secondly, the rapid transmission of SARS-CoV-2 in relation to viral evolution and the short time spans of outbreaks are insufficient for substantial genetic variation to accumulate, and therefore genomic data alone is insufficient to confer linkage or resolve the ordering of transmission; a combination of epidemiological and genomic data is required. A one SNP cut-off for defining linkage captures the majority of cases genuinely linked to the cluster, with the compromise of including a few community-onset cases likely linked by chance. The sensitivity and specificity of this SNP threshold for defining linkage varies according to the point during the pandemic at which it is being applied, both in terms of time since the start of the pandemic (greater overall viral diversity afforded by later time points in the pandemic), and the current rate of transmission (locally reduced diversity during exponential periods of spread, such as was observed with the emergence of the alpha variant in the winter of 2020). Improvements in sensitivity and specificity to detect transmission might also be gained from

considering patterns of intra-host variation[201,202]. Generally, however, joint epidemiologic and genomic analysis enable the limitations of one method to be compensated for by the strengths of the other, acknowledging both approaches are limited by undiagnosed cases.

These data have several practical IPC implications. The small proportion of cases leading to detected onward transmission highlighted that existing enhanced IPC practices were generally effective at preventing most patient-patient spread from known positive patient cases, e.g. via triaging of patients into pathways on admission and widespread diagnostic testing reducing contact between infectious/susceptible individuals. However, rates of nosocomial infection were still high at the time of this study, with the highest rates in hospitals caring for both COVID-19 and non-COVID-19 patients.

When investigating nosocomial transmission events, it is vital to consider the contributions of both patients and staff in initiating and amplifying transmission. In this study of SARS-CoV-2, the identification of superspreading events highlighted the importance of screening of both asymptomatic patients and staff to help identify and control outbreaks early (whilst acknowledging that virus in up to half of positive staff was not genetically part of the outbreak). Several strategies were applied at OUH (scaled according to the situation), including universal admission testing, weekly ward screening, asymptomatic staff lateral flow device testing as standard, scaling up to on-demand full-ward lateral flow and PCR screening, and more frequent regular screening if nosocomial cases were identified. If multiple cases are observed within a single serial interval, highly suggestive of a superspreading event, rapid action should be taken, which may involve temporary ward closure to mitigate secondary transmission, recognising that those recently infected have the highest VLs[203] and are most infectious to patients and staff[204]. If resources allow, use of dedicated staff in high risk areas, and self-isolation at home for staff exposed to a high risk

event, may also be appropriate. Challenges include recognising outbreaks spanning multiple areas and implementing effective testing and control measures, e.g. for patients who move between wards and staff who cross-cover multiple wards, including during nights and those contracted by outside agencies. Communication that patients discharged from a superspreading ward are at high risk for acquisition should lower the threshold for post-discharge SARS-CoV-2 screening/testing. Variations in rates of nosocomial transmission suggest screening should be prioritised on wards and in specialities with the highest risks (e.g. acute medicine, trauma, neurology in our setting). As vaccination-mediated reductions in inpatient COVID cases occur, it will be important to raise awareness that patients on low risk wards/pathways are still at risk of nosocomial acquisition, in addition to highlighting that in general outbreaks are caused by patients or staff not known to be positive.

Since publishing this manuscript, the COVID-19 Genomics UK Hospital-Onset COVID-19 Infection (COG-UK HOCI) trial[205], a prospective non-randomised multicentre trial to evaluate the implementation and impact of SARS-CoV-2 sequencing on IPC practice in the UK has been completed. Per protocol analysis showed that sequencing changed management in 20.7% of HOCI cases when the report was returned within 5 days, however there was no statistically change in the weekly incidence of nosocomial infection. Although the study had aimed to compare short turnaround time (<48h) to long (5-10 days), it was very challenging to obtain sequencing results within 48 hours. The median turnaround time was 5.1 days (range 3 - 29 days) across the study, most variability accounted for delay in PCR results and arrival of the sample in the sequencing lab. A shorter turnaround time was improved by increasing the proximity of the sequencing lab to the diagnostic lab, increased staff flexibility and regular transport times. A cost effectiveness analysis of the COG-UK HOCI trial showed that genomics added to the total IPC management cost[177,206].

On reflection, WGS is useful as part of a functioning system, however several conditions need to be met:

- Short turn around time, dependent on proximity/transport from clinical to research lab, reagent and consumable availability (at one point there was a shortage of pipette tips), lab staffing and sickness, which in reality is very challenging;
- Clinical team need to have the bandwidth to act on WGS results;
- Hospital has enough capacity to isolate individuals or close wards.

At points during the pandemic there was an overwhelming number of admissions, hospitals were over-capacity and there were high levels of staff illness and burnout. In these toughest periods, where hospital systems struggle to function, capacity to respond effectively to insights from WGS is hampered by the volume of cases and resource constraints[207]. In this scenario WGS is unlikely to add significantly to IPC in real-time, and time and resources may be better spent ensuring that basic IPC measures e.g. admission screening, early identification of nosocomial cases with isolation and contact tracing, are being performed well. Genomics might be better placed to retrospectively identify trends in transmissions with scheduled batch testing which can incrementally inform/improve the overall IPC strategy of the hospital.

In conclusion, I have shown that epidemiological investigation of SARS-CoV-2 infections can be enhanced by viral WGS data. I demonstrate that retrospective analyses of genomic data is useful in some circumstances to guide future IPC practice, with results consistent with similar studies in the UK[177–180]. Regardless of WGS, there is a clear need for automated systems to rapidly assimilate epidemiological data tracking patients over space and time to allow transmissions based on locations other than ward of diagnosis to be quickly identified

and fed to IPC teams. I discuss the impact of developing in-house SARS-CoV-2 sequencing further in **Chapter 7** and include a full list of SARS-Co-V-2 publications arising from the OUH HCW cohort in **Appendix 1**.

Chapter 4: Developing novel HBV whole genome sequencing methods

4.1. Availability of content

In this chapter I present findings from HBV WGS optimisation work, some of which I have published in two first author pre-prints[133,208] and presented in two first-author posters at the International HBV meeting in 2023[209,210], and a first-author poster at European Society for the Liver (EASL) conference in 2024[211]. My full manuscript is currently under review at Scientific Reports. Study sequence data have been deposited in the European Nucleotide Archive under study accession numbers PRJEB79403 for Nanopore data and PRJEB79773 for Illumina data.

4.2. Background

HBV WGS provides genetic insights into the pathogen at population and individual levels[107]. At a population level, WGS improves our understanding and monitoring of HBVDR, vaccine efficacy, treatment responses, diagnostic escape and tracking of transmission. WGS is not routinely undertaken for HBV (indeed, most infections are not even genotyped), but long term, WGS has the potential to contribute to personalised healthcare, e.g. informing modification of antiviral regimens to take into account antiviral RAMs, and supporting risk-stratified HCC surveillance[212].

The most common method for HBV sequencing is a partial-genome approach in which a short region of the HBV genome, most commonly a fragment of surface/polymerase or

pre-core/core are amplified and then sequenced with Sanger or next generation sequencing. However sequencing these isolated fragments of the genome does not give the genetic resolution required to allow accurate phylogenies to be reconstructed for sub-genotyping, identification of recombinant viruses or investigation of viral transmission. Furthermore we miss the opportunity to identify novel correlates of viral persistence, clearance, and disease, and opportunities to identify future drug and vaccine targets.

Target enrichment, the process of increasing the ratio of target to non-target nucleic acid, is required to generate HBV whole genome sequences in many samples due to the tiny HBV genome, abundant host nucleic acid background and frequently low HBV VLs < 10^4 IU/ml[108] (see **Introduction, Figure 1.10**). Traditional methods for targeted enrichment such as overlapping PCR strategies, had proven inadequate at the point of starting this work, particularly when VLs are low or when discontinuities in the rc-DNA structure complicate amplification[213]. Although single amplicon PCR methods have been developed for amplifying HBV genomes[131], they require high VLs, typically exceeding 10^4 IU/ml, which limits their applicability to patients with VLs below this threshold, who comprise a substantial proportion of the whole infected population (**Figure 1.10)**.

In this chapter I present development work performed on three viral enrichment methods (**Figure 4.1**), aiming to improve the sensitivity for generating whole genomes.

1) Rolling circle amplification (RCA)
2) Probe-based capture
3) Tiled amplicon PCR (HEP-TILE)

Scale up of WGS requires development of open access, easily deployable sequencing methods ideally building on existing infrastructure and expertise. By developing the protocols

described in this chapter, I aim to facilitate the generation of HBV sequence data globally, to obtain a more accurate and representative picture of HBV genetics.

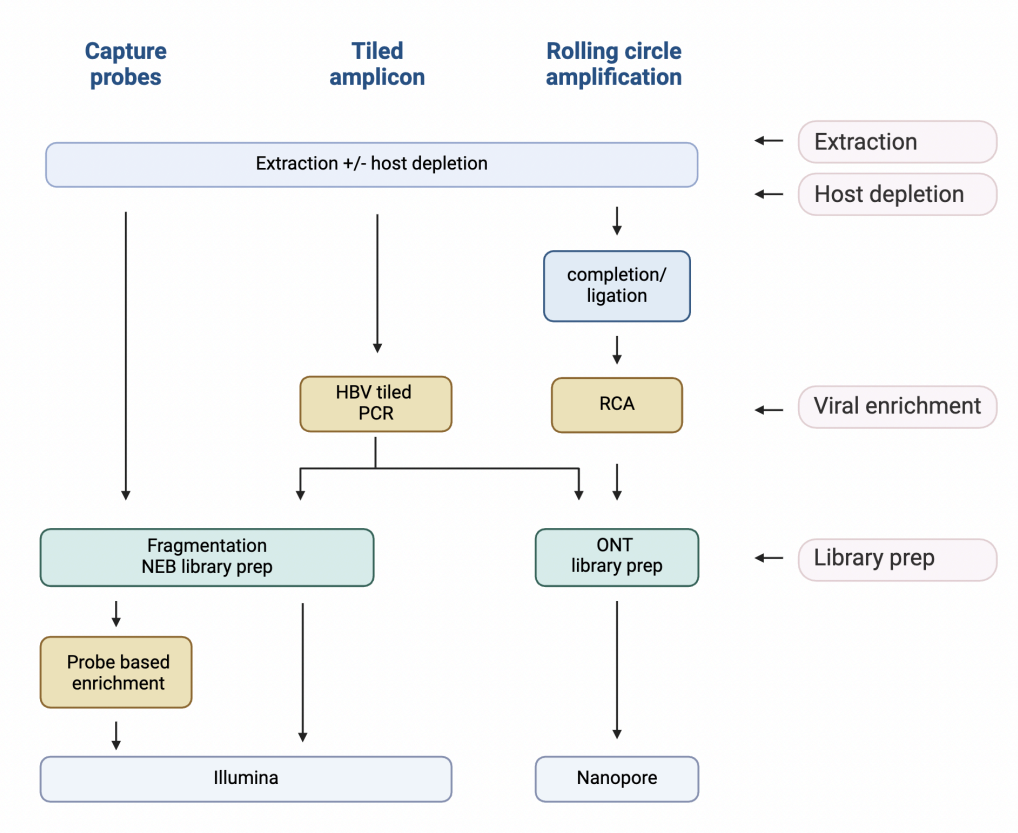


Figure 4.1. Overview of three approaches to targeted viral enrichment for HBV whole genome sequencing. Orange boxes highlight viral enrichment steps, green boxes highlight library preparation, pink boxes on the right indicate key points at which workflows could be optimised. Created in BioRender. Lumley, S. (2025) https://BioRender.com/u97g292

HBV - Hepatitis B virus; PCR - polymerase chain reaction; RCA - rolling circle amplification; ONT - Oxford Nanopore Technologies.

4.3. Aims

Aim 4.1. To improve the sensitivity of two existing methods for HBV WGS (rolling circle amplification and probe based enrichment), then due to persistent limitations in sensitivity

Aim 4.2. To develop a tiled amplicon approach (HEP-TILE) for HBV WGS

4.4. HBV rolling circle amplification

4.4.1 Background

RCA is a powerful method for the efficient amplification of circular DNA molecules. It was first introduced as a molecular biology tool in 1989 and is now predominantly performed using phi29 DNA polymerase, utilising key properties such as strand displacement activity, high processivity, and proofreading activity, which allow for the generation of long DNA products with high fidelity. In the case of circular viral genomes, RCA provides an efficient means to amplify full-length genomes. It was initially applied by Rector et al.[214] to amplify the complete circular dsDNA genome of papillomavirus.

A method for RCA to generate HBV whole genomes for sequencing was initially described by Margeridon et al. in 2008[130]. The method worked well for the fully circular cccDNA obtained from liver biopsies, however success sequencing rcDNA from serum required use of the endogenous viral polymerase to complete the rcDNA positive strand. Polymerase activity was unpredictable in samples that had been stored for long periods or undergone multiple freeze/thaw cycles. Martel et al.[144] developed an improved method of RCA amplification developing an in-vitro 'completion-ligation' step using T4 ligase.

I was part of a study in 2019 which provided proof of principle that full-genome HBV sequences can be generated by sequencing RCA-generated HBV concatemers on the Nanopore MinION (**Figure 4.2**)[143]. The bioinformatic analysis capitalised on the multi-genome concatemers to perform in-silico error correction to reduce the error rate of each genome sequenced, mitigating against the high error rate seen in ONT platforms. This study demonstrated the ability to sequence a single high VL (>8 $\log_{10}$ IU/ml) sample per ONT flow cell. Several challenges were encountered during this work including non-specific

amplification of human DNA generating hard-to-handle viscous product, and inability to successfully sequence samples with VL <6 $\log_{10}$ IU/ml. I identified several opportunities to optimise the protocol for lower VL samples at a higher throughput, the optimization work is described below:

1. Host depletion
2. Optimising the RCA reaction
3. Improving efficiency of completion ligation step
4. Enhancing sequencing of long molecules most likely to contain desired concatemers

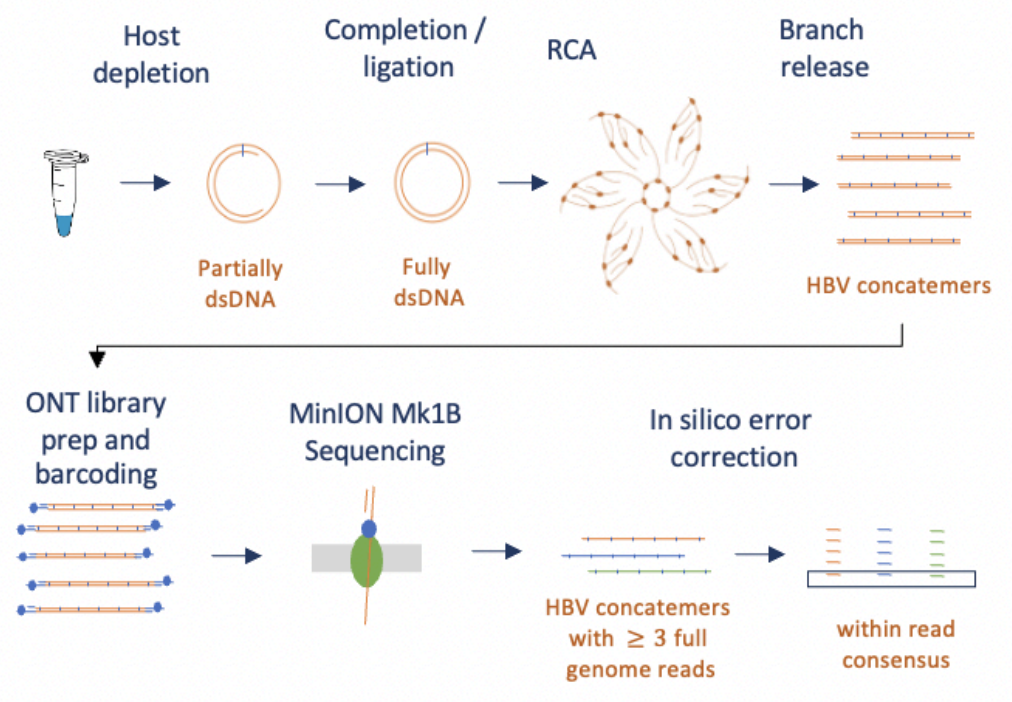


Figure 4.2. Flow diagram to show how rolling circle amplification (RCA), library preparation and bioinformatic processing can be used to generate error-corrected full length HBV genomes.

4.4.2 Methods and Results

Samples

I used the Invitrogen 1kb plus DNA ladder (catalogue number 10787018) to compare the impact of different bead clean up conditions on the DNA fragment length pre and post bead clean up. I then used clinical samples from 8 individuals to represent a range of HBV VL for my remaining work (see **Appendix 2, section 1**). Due to sample availability, I used different samples for different stages of the optimisation work.

Laboratory methods and bioinformatic protocols

See **Methods** chapter for details of primers (**Tables 2.4** and **2.5**), protocols (**Table 2.6**) and data analysis (**section 2.2.3**).

Size selection for longer DNA fragments

During analysis of RCA ONT data, all reads with < 3 full length HBV genomes are discarded, meaning that the minimum read length kept for analysis is ~10 kb long. However, the majority of DNA strands generated and sequenced during ONT library prep are <10kb in length, meaning sequence capacity is being 'wasted' on shorter reads. Using the Invitrogen 1kb plus DNA ladder, I evaluated a range of strategies for removing shorter DNA fragments, to increase the proportion of relevant length DNA that is loaded onto a flow cell:

i) SPRI bead to sample volume ratio

ii) Wash solution

iii) Elution time and temperature

iv) Sample handling

i) SPRI bead ratio

Decreasing the ratio of SPRIselect bead volume to sample volume increases the efficiency of binding larger fragments. I tested the impact of changing from a x0.6 to x0.4 clean up on a DNA ladder. The x0.4 clean up efficiently removed the majority of DNA fragments <1000bp, whereas the x0.6 cleanup only removed fragments <400bp (**Figure 4.3 A**).

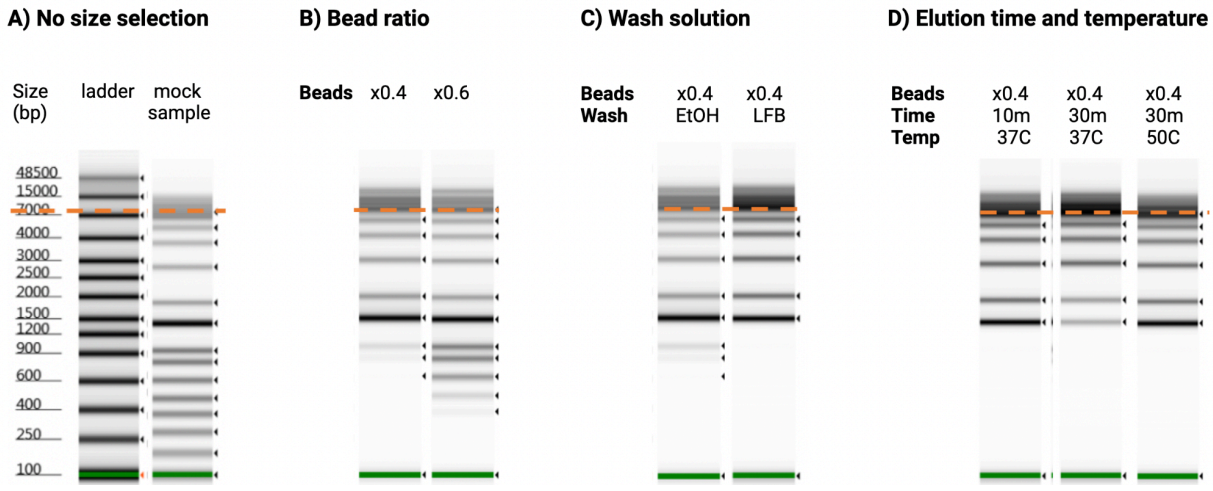


Figure 4.3. Impact of modifications to bead clean up on DNA size selection. A) DNA ladder sample without size selection, B) Comparison of x0.4 to x0.6 bead clean up, C) Comparison of x0.4 bead clean up with ethanol washes v.s. ONT long fragment buffer (LFB), D) Comparison of different elution time and temperature on the elution step of a x0.4 bead clean up with ONT LFB.

ii) Wash solution

I evaluated the impact of moving from standard ethanol washes during the bead clean up to using the ONT long fragment buffer (LFB). LFB not only eliminated DNA fragments <1200bp, but also increased the elution of fragments >1500bp (**Figure 4.3 B**).

iii) Elution time and temperature

Increasing the elution time from 10 to 30 minutes improved the concentration of long fragments relative to the shorter ones. Increasing elution temperature did not improve distribution of read lengths (**Figure 4.3 C**).

iv) Sample handling

I adopted practices initially developed to improve read length for bacteria plasmid sequencing e.g. gently flicking tubes to mix rather than pipette mixing, minimising pipetting steps where possible, where necessary pipetting gently and using the widest bore tip possible, to minimise inadvertent mechanical fragmentation of long DNA fragments during library preparation.

Host depletion

To overcome the challenge of non-specific amplification of host nucleic acid, I trialled two different methods of host depletion, aiming to selectively remove human nucleic acid prior to the RCA reaction.

(1) treatment with micrococcal nuclease (mnase), a non-specific endo-exonuclease which digests single-stranded and double-stranded DNA and RNA (depletion is performed pre-extraction), and

(2) the NEBnext microbiome DNA enrichment kit, which removes CpG-methylated host DNA using magnetic beads (depletion is performed post-extraction).

I investigated whether the host depletion method improves sequencing output. I took four samples (S1-4) at a range of viral loads (7.47, 5.69, 3.40 and 1.65 $\log_{10}$ IU/ml) (**Appendix 2, section 1**) and compared the impact of no depletion v.s. mnase v.s. NEBnext microbiome kit (see **Methods, section 2.2.2**) on the percentage of HBV and human reads and the absolute number of HBV reads generated. I used equivolume pooling post-barcoding (rather than equimolar balancing) to allow comparison of methods. Due to sample availability I was not able to perform the experiment in duplicate/triplicate.

Although the NEB microbiome kit reduced the percentage of human reads most effectively in 3 out of 4 samples (**Figure 4.4 A**), this did not consistently translate into an increase in the number of HBV reads (**Figure 4.4 B**). Host depletion led to improvements in the number of HBV reads in S1 and S2, however had minimal impact on samples S3 and S4 with lower VL (**Figure 4.4 B**). Only the sample with the highest VL (S1) generated reads containing >3 HBV concatemers; the S1 replicate without host depletion generated 14 reads with >3 concatemers, the replicate treated with mnase generated 695 reads with >3 concatemers and the replicate treated with the NEB microbiome kit generated 1022 reads with >3 concatemers. Given the differences in cost and time required for the methods (30 minutes and ~ £1 per sample for mnase and 90 mins and ~£40 per sample for NEB microbiome kit, I used mnase in future experiments.

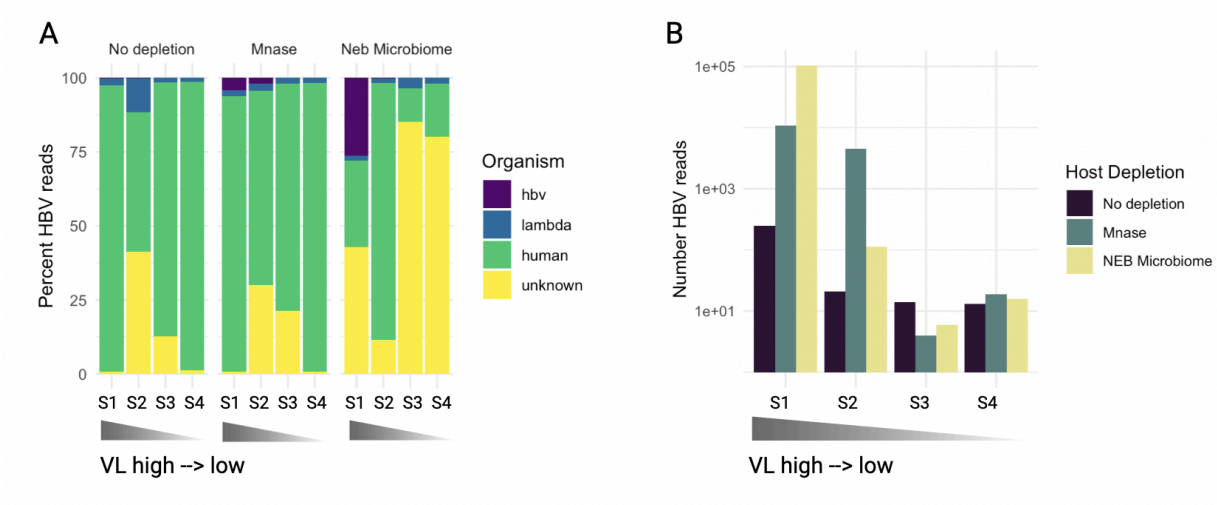


Figure 4.4. Impact of host depletion on HBV sequencing outputs in samples enriched by rolling circle amplification. A) Impact of host depletion on the percentage of reads mapping to the HBV genome. B) Impact of host depletion on the number of HBV reads. VL of samples is indicated by black shaded bar, S1 (highest VL) to S4 (lowest VL).

RCA reaction and debranching

RCA has historically been performed using the phi29 DNA polymerase, a highly processive polymerase with strong strand displacement activity which allows for efficient isothermal DNA

amplification. The end product is hyper-branched with multiple concatenated copies of the original target, and a combination of single stranded and double stranded regions. The use of multiple sets of forward and reverse primers (eight in the original protocol[130]) leads to more efficient amplification as the primer can bind to the displaced DNA strand and initiate additional DNA synthesis. It also leads to conversion of some ssDNA to dsDNA, which is required for library preparation and sequencing. The more primer sites, the more hyperbranched the structure is, which leads to challenges with handling of the viscous product. Furthermore, theoretically more primer sites also lead to shorter dsDNA stretches with fewer concatemers on each branch. Due to the continuous binding of primers and generation of new branches, some DNA will remain single stranded.

To try to overcome these challenges I trialled three method modifications:

(1) Switching from phi29 polymerase to equiphi29 polymerase - capitalising on improved reaction speed and increased dsDNA yield to generate more usable dsDNA in a shorter timeframe (2 vs. 20 hours)
(2) Reducing the number of primer pairs - investigating the trade off between amplification dynamics and product length
(3) Introducing a debranching step to increase the of free dsDNA ends suitable for library preparation

Firstly, I investigated the ability of phi29/equiphi29 with/without debranching to generate HBV concatemers on the ONT MinION. I used four samples (S1-4) which I treated with mnase to deplete host DNA prior to extraction on the kingfisher apex. I performed phi29 RCA using the Martel et al. protocol[144] and Equiphi29 RCA as per manufacturer's recommendations followed by debranching (see **Methods chapter, Table 2.6**).

Equiphi RCA led to 100 fold less enrichment than phi29 (determined by HBV qPCR performed pre and post RCA). Only S1, the highest VL sample, consistently generated reads with more than 3 HBV concatenated genomes; phi29 RCA generated ~4000 reads with >3 concatenated genomes, whereas equiphi generated ~200 reads. Debranching did not impact on the number of reads with >3 HBV concatemers (**Figure 4.5**).

Next I investigated the impact of primer number by performing RCA on two samples (S5/6, (Appendix 2, section 1) 3.9 and 5.00 $\log_{10}$ IU/ml), comparing the effect of using all 8 primers, compared to a pool containing only 6 (primers RCA 1-4, 7 and 8 Methods chapter table 2.4) or 4 primers (primers RCA 1-4). My hypothesis was that reducing the number of primers would reduce in a product that had fewer but longer branches, as enzyme activity would be focused on extending fewer start sites for a longer duration. I analysed the total DNA yield using Qubit, the number of HBV copies using HBV qPCR, and the length of the product using Agilent Tapestation genomic DNA screentape. As expected, reducing the number of primers in the primer pool reduced the total DNA yield and the fold enrichment of HBV determined by qPCR, however it unexpectedly decreased the mass of product that was >10kb in length (**Figure 4.5**). In summary, the optimisations to the RCA reaction did not improve the mass of long dsDNA product or permit the generation of HBV concatemers from samples < 6 $\log_{10}$ IU/ml.

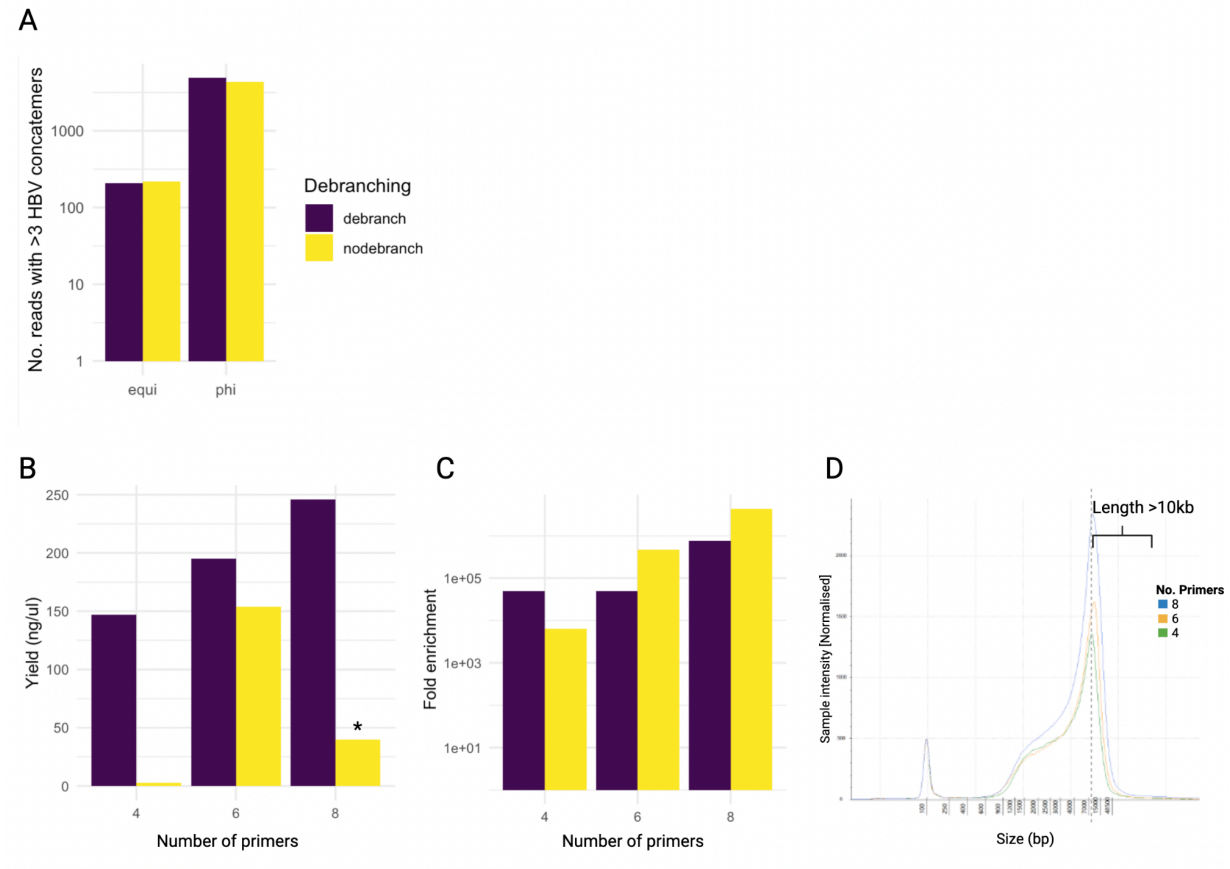


Figure 4.5. Optimisation of rolling circle amplification protocol enzymes and primers.

A) Impact of changing enzyme from phi to equiphi and introducing a de-branching step on the number of reads containing HBV concatemers with >3 full length copies of the HBV genome. B) Impact of decreasing primer number on total DNA yield of RCA. C) Impact of decreasing primer number on fold HBV enrichment determined by HBV qPCR. D) Impact of decreasing primer number on length of total DNA product produced by RCA using Tapestation.

* indicates hyperviscous sample challenging to quantify using Qubit

Improving generation of circular template

In the presence of a circular template, RCA should amplify a circular product extremely efficiently. I hypothesised that the lack of amplification of full length HBV genomes in all but the highest VL samples, observed by myself and elsewhere[140,141,215], might be due to the lack of a fully circular template. Either the completion ligation (CL) reaction, which converts rcDNA to cccDNA was very inefficient, only converting rcDNA to a fully circular template in the

highest VL samples, or RCA was independent of the CL reaction, potentially as a result of the presence of an alternative circular template (i.e. cccDNA) in the plasma of individuals with very high VL. There is also a risk that the HBV genomes were fragmented in some samples.

HBV rcDNA has HBV polymerase covalently bound to the negative strand (**Figure 4.6 A**). The removal of this polymerase is required to 'repair' the rcDNA to a fully continuous double stranded molecule. Previous studies mimicking *in vivo* cccDNA with an artificial HBV construct highlighted that the presence of a protein adduct slowed minus strand repair which was the rate-limiting step for cccDNA formation[216]. Although the extraction methods I use have a protease step, this does not completely remove the HBV polymerase, but leaves a short remnant peptide covalently attached to the minus strand flap[217]. Therefore I hypothesised that the polymerase/remnant peptide might be inhibiting efficient circularisation of the rcDNA template.

I identified three novel methods for generating a circular template for RCA (**Figure 4.6 B,C,D**). The first aimed to mimic in vivo cccDNA formation by using flap endonuclease 1 (FEN1) to remove the flap/DNA adduct in combination with a strand displacing polymerase (equiphi chosen over phi29 due to reaction temperature being closer to temperature for optimum ligase activity) and Hi-T4 DNA ligase for ligation (**Figure 4.6 B**). Two further options included performing a full length PCR, running the PCR product on a gel, cutting out and extracting the band corresponding to a full length HBV product, digesting then ligating the full length product with a ligase to form a dsDNA circular template (**Figure 4.6 C**), or taking the full length product and ligating hairpin adaptors to form a ssDNA circular template (**Figure 4.6 D**). I describe the protocols in **Methods chapter, Table 2.6.**

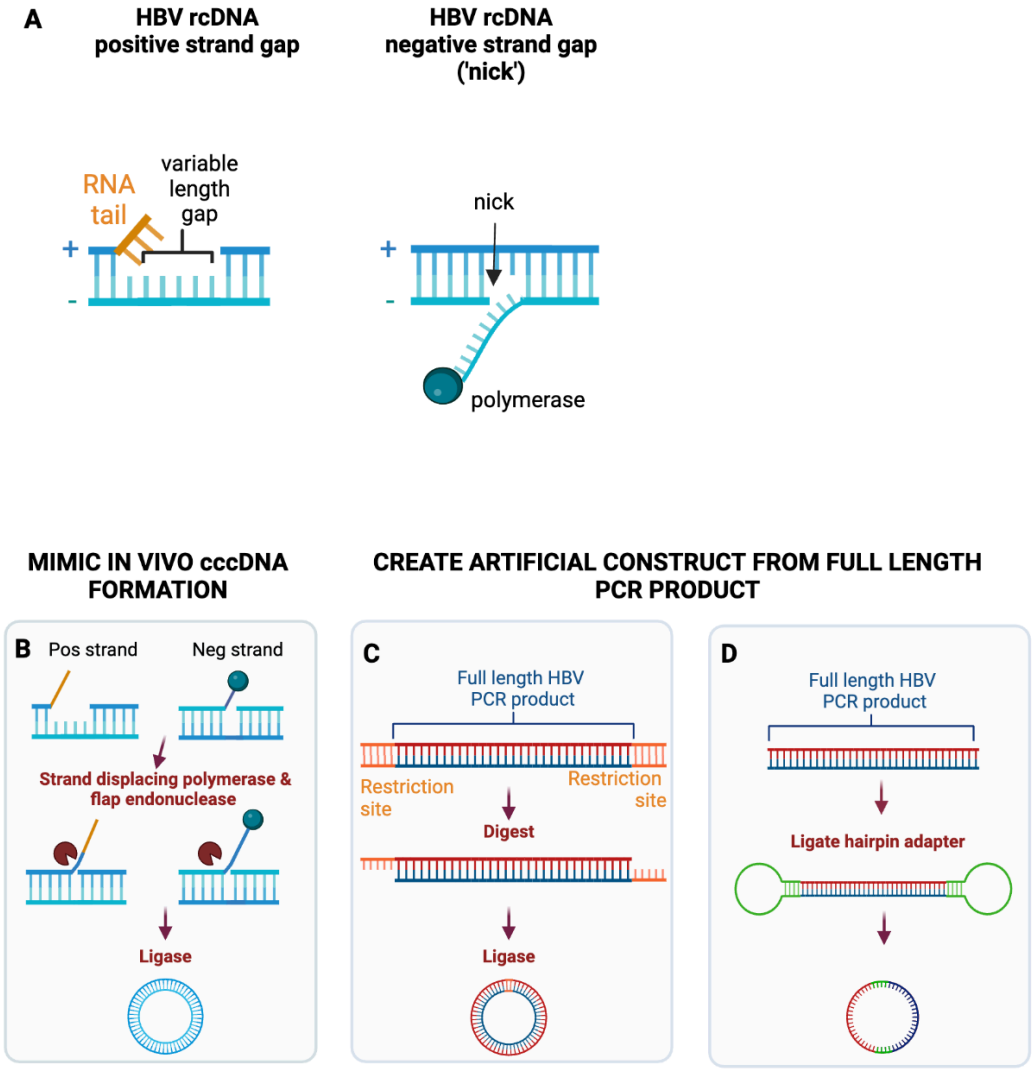


Figure 4.6. Strategies for generating a fully circular HBV template for rolling circle amplification. A) HBV rcDNA structure showing gaps in positive and negative trends; B) Mimic in vivo cccDNA formation with a strand displacing polymerase and flap endonuclease to lift and remove the DNA flap with bound polymerase; C) Generate a full length HBV amplicon using PCR primers with restriction enzyme site. Use restriction enzyme to generate sticky ends and ligase to ligate into a double stranded circular template; D) Generate a full length HBV amplicon, ligate hairpin adaptors to each end to generate a single stranded circular template (with 2 genome copies). Created in BioRender. Lumley, S. (2025) https://BioRender.com/z11j329

I used samples S7/8 to trial gPCR-RCA and the hairpin adaptor and S9/10 to trial FEN1-CL (see **Appendix 2** section 1, NB this was not a direct head to head of methods). I was able to successfully generate concatenated HBV genomes using the full genome PCR followed by RCA (gPCR-RCA); S7 with VL 8.23 $\log_{10}$ IU/ml generated 101 reads with ≥3 full HBV genomes concatenated, S8 with VL 5.50 $\log_{10}$ IU/ml generated 37221 reads with ≥3 full HBV genomes concatenated.

No reads with ≥3 concatenated genomes were generated with FEN1 completion/ligation (samples S9 and S10, **Appendix 2**, section 1). Quantification, post RCA, of S7/8 PCR product ligated with the hairpin adaptor was low, so I did not take this forward for sequencing.

4.4.3 Conclusion

RCA with ONT sequencing and in-silico error correction is an appealing method for HBV WGS due to the ability to confidently reconstruct the whole HBV haplotype and correct errors in silico. However despite trying extensive optimisation, I was unable to improve sensitivity using the most straightforward approach of completion/ligation followed by RCA. I hypothesised that this is due to the lack of an appropriate circular template, likely to be a combination of an inefficient completion/ligation step and potential viral genome degradation at low VL.

I successfully improved RCA sensitivity by moving to a “gPCR-RCA” approach, adding a full length genomic PCR and circularisation of the PCR product prior to the RCA reaction. However, although achievable, this approach has limitations. The limiting step of the gPCR-RCA workflow is the gPCR, which still requires the presence of intact viral rcDNA which can be degraded in low VL samples. Although gPCR-RCA may be able to improve

slightly on the sensitivity of gPCR by providing additional amplification of the PCR product, it is unlikely to substantially improve sensitivity for samples with VL < 3 $\log_{10}$ IU/ml, which has been the previous limit for full length gPCR. Furthermore, each incremental step added to an enrichment protocol not only increases time and cost of the protocol, but also the risk of introducing PCR errors or cross-contamination of samples. Due to this combination of factors, I did not perform any further validation work comparing RCA with/without gPCR or test the limit of detection for gPCR-RCA, and instead moved on to developing other methods for HBV WGS.

4.5. Probe-based enrichment

4.5.1 Background

Due to the limitations of the RCA protocol described above, I moved to probe-based enrichment, a strategy which has previously been validated across diverse pathogens, including by collaborators at Oxford University (Ansari et al) for sequencing HCV [134,218,219]. Here I describe the design, optimisation and use of a pan-genotypic panel of probes to improve sequencing yield for HBV. This work has been published in an early pre-print published to make the method available in the public domain, followed by a later update which is now under peer review[133,208].

4.5.2 Methods

I used 20 clinical samples from adults with CHB with VL < ~4 $\log_{10}$ IU/ml, and 23 clinical samples with VL > ~4 $\log_{10}$ IU/ml (see **Appendix 2, section 2** and **Figure 4.7** for sample details). I prepared mock HBV samples representing genotypes A-J (see **Methods** chapter, section 2.2.1, for rationale and details). The wet lab workflows and bioinformatic pipelines used for different phases of the projects are described in **Chapter 2**.

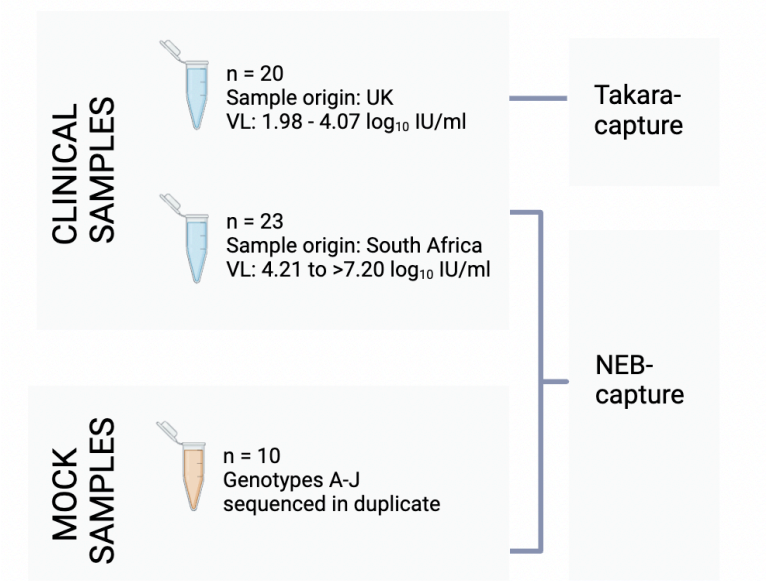


Figure 4.7. Flow chart summarising sample sets for enrichment with probe based capture. I used an initial set of 20 clinical samples from the UK to evaluate Takara library prep followed by capture ("takara-capture") and a 2nd set of 23 clinical samples from South Africa alongside 10 mock samples, to evaluate New England Biolabs (NEB) library prep followed by capture ("NEB-capture") (see **Methods** chapter, **Table 2.6** for further details of methods). Created in BioRender. Lumley, S. (2025) https://BioRender.com/p12e458

4.5.3 Results

Probe-based enrichment increases HBV sequence yield for samples < 4 log_{10} IU/ml

I compared the sequencing output of pooled libraries of 20 samples (**Figure 4.7, Appendix 2 section 2**) with/without HBV specific probe-based enrichment (termed "capture" vs "no-capture" (ie. metagenomic) respectively) using takara library prep (see **Chapter 2, Table 2.6**). I selected samples with HBV VL < ~4 log_{10} IU/ml, representing harder to sequence low VLs.

Capture increased the number of HBV reads per million reads sequenced (rpm), resulting in a 47,970 fold increase in the number of reads mapped to the HBV genome (243 HBV rpm in

the capture pool vs. 5×10^{-3} rpm in the no-capture pool). Out of 20 samples, only 1/20 without capture generated HBV reads (VL 3.89 $\log_{10}$ IU/ml); this increased to 19/20 samples with capture. HBV sequence yield was increased in the capture arm and resulted in 2.30 $\log_{10}$ (95% confidence interval 1.99 - 2.48 $\log_{10}$) increase in HBV rpm per $\log_{10}$ increase in VL (**Figure 4.8**) when a linear relationship was assumed. Capture increased the proportion of HBV reads from a median of 0 (range 0 - 8.05 $\times10^{-8}$) to a median of 3.75 $\times10^{-5}$ (range 0 - 6.17×10^{-4}) whilst decreasing the proportion of human reads. However, despite the enrichment in number of HBV reads, <0.1% reads mapped to HBV, only partial genomes could be assembled and median depth of coverage remained 0 (**Table 4.1**), I concluded that capture was unlikely to be suitable for samples of VL < 4 $\log_{10}$ IU/ml and moved on to working with VLs > 4 $\log_{10}$ IU/ml.

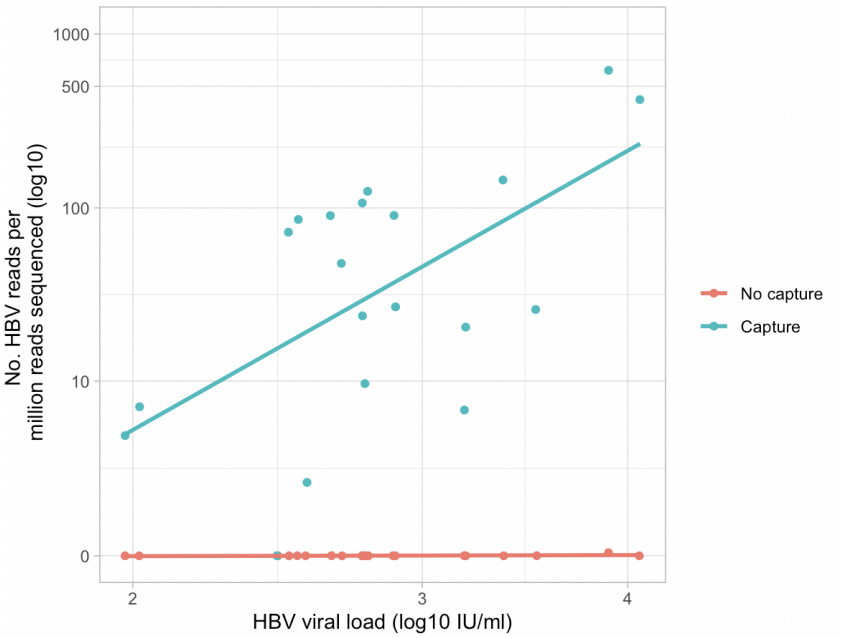


Figure 4.8. Comparison of number of HBV reads per million sequenced with/without probe-based capture. HBV specific probe-based capture increased the efficiency of HBV sequencing by nearly 48,000 fold. Assuming a linear relationship between sequence yield and VL, I observed a 2.30 $\log_{10}$ increase in the number of HBV reads per $\log_{10}$ increase in HBV DNA VL in plasma.

Table 4.1. Viral load and sequencing output for 20 plasma samples from adults with CHB enriched using 'Takara-capture' protocol and sequenced by Illumina.

Sample ID	VL (log10 IU/ml)	No capture						Capture					
		Read count	Reads pass-QC	HBV reads	HBV reads per million	% HBV genome assembled	Median coverage	Read count	Reads pass-QC	HBV reads	HBV rpm	% HBV genome assembled	Median coverage
HEP 1711	4.07	20933044	17463756	0	0	0	0	15053030	12226942	6322	420	95	9
HEP 1546	3.89	24829796	20747830	2	0.1	0	0	17997266	14653472	11111	617	96	10
HEP 1452	3.52	21145796	17416474	0	0	0	0	15114326	12080868	393	26	48	0
HEP 1522	3.36	20230320	17013454	0	0	0	0	14408928	11803456	2085	145	88	3
HEP 1489	3.19	25862232	21354586	0	0	0	0	19474444	15652348	401	21	48	0
HEP 1191	3.18	19398596	16257044	0	0	0	0	13052120	10503632	88	7	0	0
HEP 1385	2.89	21290168	18053908	0	0	0	0	15201562	12546846	410	27	26	0
HEP 1028	2.88	17752464	14917026	0	0	0	0	12927836	10557494	1171	91	59	1
HEP 1018	2.78	12586968	9477256	0	0	0	0	9346712	6768854	1163	124	55	1
HEP 1462	2.77	23567564	17763880	0	0	0	0	17390788	12429772	168	10	34	0
HEP 1165	2.76	14323558	12356090	0	0	0	0	9752656	8196682	1041	107	0	0
HEP 1401	2.76	18172348	15403362	0	0	0	0	12879278	10591494	308	24	0	0
HEP 1622	2.68	23259426	19678708	0	0	0	0	16773174	13860314	805	48	73	1
HEP 1328	2.64	19579800	15530932	0	0	0	0	13960764	10578,95	1261	90	54	1
HEP 1190	2.55	18320118	14323370	0	0	0	0	13272736	9921674	30	2	0	0
HEP 1341	2.52	17431576	14992498	0	0	0	0	12095246	10120378	1037	86	35	0
HEP 1433	2.49	13876540	10698164	0	0	0	0	9820456	7184016	712	73	0	0
HEP 1167	2.45	20962554	17802228	0	0	0	0	15212438	12532	0	0	0	0
HEP 1099	2.02	22997532	18572882	0	0	0	0	17012832	13351474	120	7	0	0
HEP 1149	1.98	18542206	15610616	0	0	0	0	13399040	11001066	63	5	12	0
Median		**19905060**	**16635249**	**0**	**0**	**0**	**0**	**14184846**	**11001066**	**561**	**38**	**35**	**0**

QC - Quality control; VL - viral load

Probe-based enrichment generates full genomes for samples > 4 log_{10} IU/ml

I enriched a set of 23 clinical samples with higher VL 4.21 to >7.20 log_{10} IU/ml (**Figure 4.7**, details in **Appendix 2 section 3**) using probe based capture (NEB library prep, see **Chapter 2, Table 2.6**) and sequenced the samples on Illumina Novaseq X. A median of 22,377 (IQR 3267 - 135,184) HBV rpm was generated. The relationship between HBV VL and percentage reads on target is shown in (**Figure 4.9**). Full length (>98%) genomes at x1 coverage were generated in 18/23 samples, with partial genomes in 5/23. Full length genomes above a minimum x5 read depth (minimum required by bioinformatic pipeline to call base at the consensus level) were generated in 7/23 samples, all of which had VL >5.5 log_{10} IU/ml (~300,000 IU/ml). Below VL 5.5 log_{10} only partial genomes at x5 depth were generated (**Table 4.2**). The negative controls had 0% consensus coverage of the HBV genome.

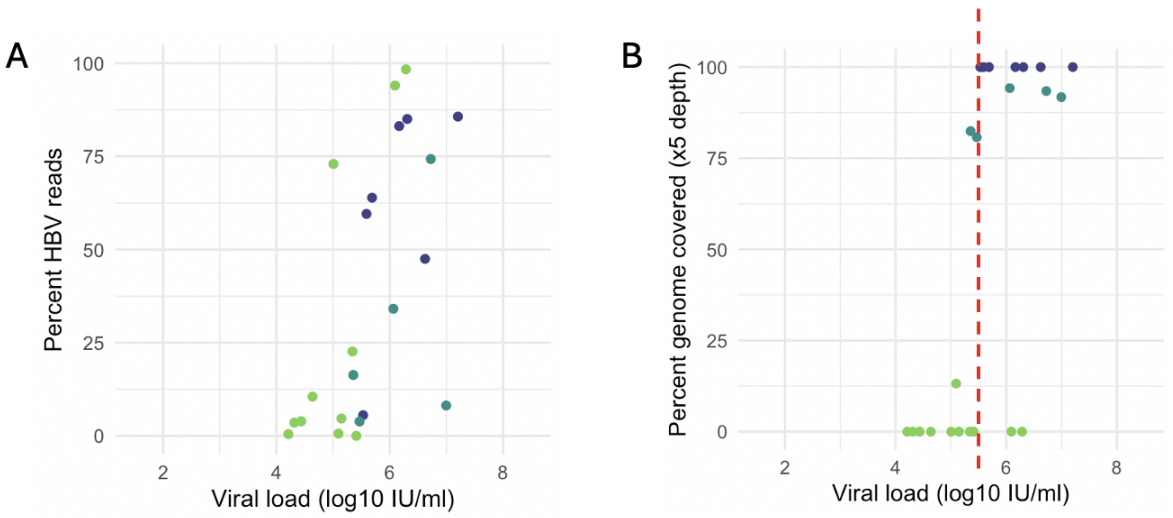


Figure 4.9. Sequencing outputs for HBV WGS using capture for viral enrichment.

A) Relationship between VL and percentage of reads on-target (ie. HBV) for probe-based capture. B) Relationship between VL and percentage of genome with high confidence consensus for probe-based capture. Red dotted line shows the threshold (5.5 log_{10} IU/ml) above which capture can generate full HBV genomes.

Table 4.2. Sequencing outputs for clinical HBV samples enriched by NEB-capture.

Sample ID	HBV VL (log10 IU/ml)	Genotype	Capture Illumina					
			Total reads	Reads mapped to HBV	HBV reads per million sequenced	Percent reads mapped to HBV	Percent genome coverage (x1)	Percent genome coverage (x5)
1113	> 7.2	A	4177836	3579271	368492	36.85	100.00	100.00
1162	7.00	A	3165736	257935	32931	3.29	100.00	91.76
1143	6.73	A	513650	381579	287947	28.79	100.00	93.42
1191	6.62	A	4987814	2368758	196465	19.65	100.00	100.00
1114	6.31	A	2219342	1887048	339437	33.94	100.00	100.00
251	6.29	A	31276	30765	1727	0.17	97.61	0.00
1125	6.17	A	638778	530971	114791	11.48	100.00	100.00
200	6.09	A	109738	103162	14252	1.43	99.81	0.00
1175	6.06	A	1182264	403223	127007	12.70	100.00	94.23
1103	5.69	A	966700	617788	143360	14.34	100.00	100.00
218	5.59	A	1708810	1017820	256890	25.69	100.00	100.00
1119	5.53	A	5746052	318907	1577	0.16	100.00	100.00
1131	5.47	A	3851170	148765	6228	0.62	100.00	80.75
1116	5.41	A	2958010	399	2	0.00	75.28	0.00
1105	5.36	A	1057854	172593	80165	8.02	100.00	82.43
1115	5.34	A	434768	98510	40624	4.06	100.00	0.00
1132	5.15	not done	646084	30054	3984	0.40	99.53	0.00
1147	5.09	not done	5898812	36612	342	0.03	100.00	13.19
211	5.01	not done	43616	31822	22377	2.24	100.00	0.00
1108	4.64	not done	459994	48425	3270	0.33	100.00	0.00
1174	4.44	not done	355378	13899	3264	0.33	91.34	0.00
1134	4.32	not done	1028776	36627	5168	0.52	95.19	0.00
1170	4.21	not done	5590660	26066	151	0.02	91.52	0.00
Median			**1057854**	**148765**	**22377**	**2.24**	**100.00**	**80.75**

Sequencing mock HBV samples enriched with probe based capture using Illumina

In order to verify the pan-genotypic performance of the probes, I sequenced 10 mock samples (genotypes A-J, see **Methods** chapter) in duplicate, with/without probe-based capture on the Illumina Novaseq X. I diluted libraries to 10pM and pooled prior to sequencing. Without capture, a median of 391 (IQR 153 - 1858) HBV reads per million reads sequenced was generated across 20 mock samples. With capture a median of 563,956 (IQR 397523 - 896635) HBV reads per million reads sequenced was generated (**Table 4.3**). Enrichment with probe-based capture led to the generation of full length HBV consensus sequences for all genotypes (**Figure 4.10, Table 4.3**).

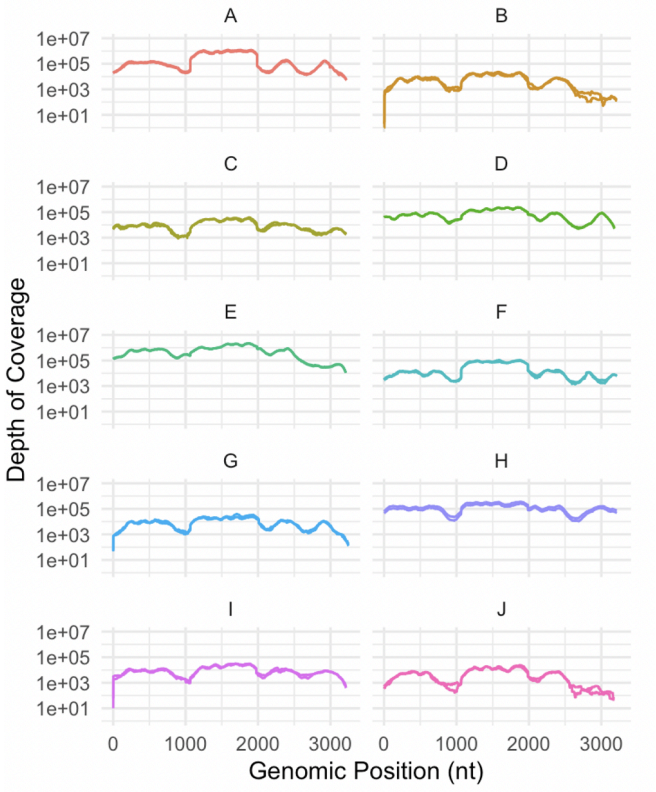


Figure 4.10. HBV genome coverage for mock samples with HBV genotypes (A-J) enriched using HBV probe based capture.

Table 4.3. Sequencing metrics for mock HBV samples prepared with/without probe based capture and sequenced on Illumina.

Genotype	Reference genome	No enrichment - illumina sequencing					Capture Illumina					
		Total reads	Mapped reads	HBV reads per million sequenced	Percent reads mapped to HBV	Percent genome coverage	Total reads	Mapped reads	HBV reads per million sequenced	Percent reads mapped to HBV	Percent genome coverage (x1)	Percent genome coverage (x5)
A	KU736919	2564112	10405	4058	0.41	100	7703618	6946222	901683	90.17	100	100
		2135330	7887	3694	0.37	100	9257976	8358193	902810	90.28	100	100
B	not published	2545038	330	130	0.01	0	511478	149377	292050	29.20	100	100
		2030784	314	155	0.02	0	551530	172912	313513	31.35	100	100
C	MT111596	0	0	0	0.00	0	727296	309195	425130	42.51	100	100
		2819740	449	159	0.02	0	667198	270331	405174	40.52	100	100
D	NC_003977	2594796	3433	1323	0.13	100	2359894	1997417	846401	84.64	100	100
		1787174	2604	1457	0.15	100	2365408	1999929	845490	84.55	100	100
E	KU736895	2456130	42317	17229	1.72	100	17116198	16524062	965405	96.54	100	100
		2050518	36693	17895	1.79	100	17556920	16941576	964951	96.50	100	100
F	DQ823095	2446438	1418	580	0.06	98.48	1069550	776761	726250	72.63	100	100
		1355114	750	553	0.06	83.05	981698	671782	684306	68.43	100	100
G	AB056513	2390022	357	149	0.01	0	605588	226835	374570	37.46	100	100
		2083668	334	160	0.02	0	708842	290037	409170	40.92	100	100
H	FJ356716	2382670	4789	2010	0.20	100	3247752	2906583	894952	89.50	100	100
		1250400	2260	1807	0.18	100	4593134	4154349	904469	90.45	100	100
I	AB562463	2476292	530	214	0.02	0	612856	270695	441694	44.17	100	100
		1342546	308	229	0.02	0	538454	238861	443605	44.36	100	100
J	AB486012	2729238	357	131	0.01	0	470730	147289	312895	31.29	100	100
		1779390	248	139	0.01	0	474726	129406	272591	27.26	100	100
Median		**2259000**	**640**	**391**	**0.04**	**41.525**	**854497**	**490489**	**563956**	**56.40**	**100**	**100**

4.5.4 Conclusion

Although our capture approach can be used to improve the efficiency of viral sequencing at VL >~5.50 $\log_{10}$ IU/ml, its efficacy for producing full-length consensus genomes diminishes at lower VLs, as observed elsewhere[220]. Additional/alternative host depletion and high volume (e.g. 5ml[127]) extraction is likely to be required to reliably obtain full consensus genomes using this method in samples with VL <5.5 $\log_{10}$ IU/ml (shown as red dotted line in **Figure 4.9 B**). Above this threshold 10/12 (83%) had >90% genome coverage, whereas below this threshold no samples generated >90% coverage. An alternative method is required to sequence samples with VL <5.5 $\log_{10}$ IU/ml.

4.6. HEP-TILE: tiled amplicon HBV WGS

4.6.1 Background

Due to limitations in sensitivity of RCA, probe-based enrichment, and full genome PCR schemes, I developed a tiled amplicon approach for HBV whole genome sequencing (HEP-TILE). The HBV sequencing wet and dry lab workflows build on the widely used ARTIC-Network nCoV-2019 sequencing protocol[148] and fieldbioinformatics[221] respectively, both of which have been used extensively worldwide during the COVID-19 pandemic and which I used in the first half of my D.Phil. to perform SARS-CoV-2 WGS.

4.6.2 Methods

I used 50 clinical samples from adults with CHB to test the HEP-TILE Nanopore pipeline, 10 of which I also used to trial the HEP-TILE Illumina pipeline (see **Appendix 2, section 4** and **Figure 4.11** for sample details). I also prepared mock HBV samples representing genotypes A-J (see **Chapter 2** for rationale and details). I have described the primer scheme development, wet lab pipeline and bioinformatic pipeline in detail in the **Chapter 2, section 2.2.2**.

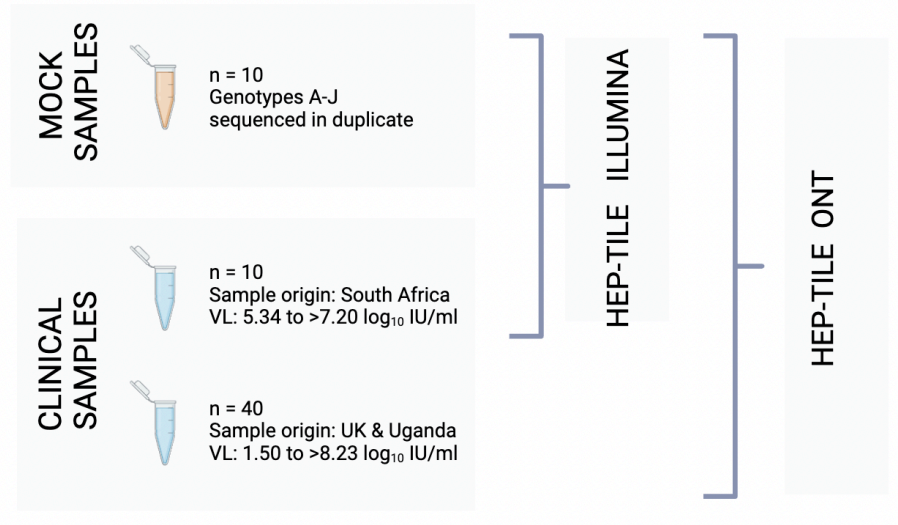


Figure 4.11. Sample sets for tiled amplicon (HEP-TILE) enrichment optimisation.

All samples were sequenced with Nanopore, residual sample was available for mock samples and clinical samples from South Africa for sequencing on Illumina. Created in BioRender. Lumley, S. (2025) https://BioRender.com/p06t038. ONT - Oxford Nanopore Technologies.

4.6.3 Results

Sequencing clinical HBV samples enriched with HEP-TILE using Nanopore

I enriched a subset of 50 clinical samples (HBV VL 1.50 to >8.23 log_{10} IU/ml, **Figure 4.11, Appendix 2 section 4**) using the HEP-TILE protocol and sequenced on the Nanopore MinION. I generated a median of 779,909 (IQR 590,906 - 676,536) HBV rpm (**Table 4.4**). The relationship between HBV VL and percentage reads on target is presented (**Figure 4.12 A and C**).

Table 4.4. Sequencing outputs for clinical HBV samples enriched by HEP-TILE.

Sample ID	Country	HBV VL (log10 IU/ml)	Genotype	HEP-TILE Nanopore						HEP-TILE Illumina					
				Total reads	Reads mapped to HBV	HBV reads per million sequenced	Percent reads mapped to HBV	Percent genome coverage (x1)	Percent genome coverage (x20)	Total reads	Reads mapped to HBV	HBV reads per million sequenced	Percent reads mapped to HBV	Percent genome coverage (x1)	Percent genome coverage (x5)
1113	South Africa	> 7.2	A	32085	18893	588842	58.88	100	99.94	2799578	2740035	978731	97.87	100	100
1143	South Africa	6.73	A	37594	25733	684498	68.45	100	100	3194506	3131012	980124	98.01	100	100
1191	South Africa	6.62	A	71800	53081	739290	73.93	100	100	4300788	4115008	956803	95.68	100	100
251	South Africa	6.29	A	31326	20445	652653	65.27	100	99.69	2895860	2838728	980271	98.03	98.79	98.7
200	South Africa	6.09	A	60856	18960	311555	31.16	83.20	83.17	1767254	582508	329612	32.96	98.1	84.8
1175	South Africa	6.06	A	62380	38099	610757	61.08	100.00	100.00	3268324	2744378	839690	83.97	100	100
218	South Africa	5.59	A	117575	90275	767808	76.78	100.00	99.91	4099128	3922030	956796	95.68	100	100
1119	South Africa	5.53	A	95703	35680	372820	37.28	100	100	4276904	3025569	707420	70.74	100	100
1116	South Africa	5.41	A	187499	31	165	0.02	0	0	1596600	1169135	732265	73.23	0	0
1115	South Africa	5.34	A	19235	4298	223447	22.34	37.75	37.72	2420162	2359125	974780	97.48	67.26	59.9
HEP-1363	England	5.64	B/C	61511	59051	960007	96	100	100						
HEP-1745	England	5.5	D/E	49488	43628	881587	88.16	100	99.59						
HEP-1740	England	4.82	B/C	77987	76054	975214	97.52	100	99.53						
HEP-0709	England	4.51	B/C	9329	3982	426841	42.68	100	99.66						
HEP-1749	England	4.49	E	33654	26561	789238	78.92	100	99.72						
HEP-1740	England	4.4	B/C	1189	490	412111	41.21	100	99.6						
HEP-1727	England	4.1	C	15312	9030	589734	58.97	100	100						
HEP-1763	England	3.79	A	25568	6657	260365	26.04	100	99.69						
HEP-1745	England	3.73	D/E	61411	45255	736920	73.69	100	99.5						
HEP-1747	England	3.35	D	6324	757	119703	11.97	84.63	84.32						
HEP-1571	England	3.16	A	3494	1115	319118	31.91	100	99.5						
HEP-1403	England	3.01	A/D	6140	1898	309121	30.91	85.01	84.57						
HEP-1104	England	2.63	D/E	86946	65473	753031	75.3	100	99.81						
HEP-1752	England	2.55	C	476372	436221	915715	91.57	100	99.56						
HEP-1746	England	2.52	D/E	216978	63134	290970	29.1	100	99.94						
HEP-1766	England	2.18	D	78875	43793	555220	55.52	100	99.87						
HEP-1681	England	1.9	B/C	4717	1342	284503	28.45	100	99.88						
HEP-1529	England	1.62	D	46990	260	5533	0.55	35.54	35.54						
HEP-1710	England	1.5	E	81646	24171	296046	29.6	100	99.81						
900219N	Uganda	8.23	A	1461622	1460216	999038	99.9	100	99.97						
906667N	Uganda	8.23	A	1644431	1634360	993876	99.39	100	99.97						
907906N	Uganda	7.6	A	853932	853530	999529	99.95	100	99.91						
900603N	Uganda	7.5	A	1214437	1212327	998263	99.83	100	99.97						
906116N	Uganda	7.02	A	850846	830695	976317	97.63	100	99.75						
ULIDS-0973	Uganda	6.13	A	1183724	1180867	997586	99.76	100	99.72						
ULIDS-0610	Uganda	4.64	A	487594	486034	996801	99.68	100	99.97						
907395N	Uganda	4.05	A	155913	154865	993278	99.33	84.97	84.82						
901116N	Uganda	3.9	A	233603	227036	971888	97.19	100	99.91						
ULIDS-0434	Uganda	3.72	A	219108	195817	893701	89.37	87.92	87.71						
ULIDS-0001	Uganda	3.63	A	814736	811945	996574	99.66	100	99.29						
ULIDS-0433	Uganda	3.61	A	93667	77027	822349	82.23	84.63	84.1						
903818N	Uganda	3.53	A	105909	100711	950920	95.09	100	99.75						
ULIDS-0457	Uganda	3.36	A	188262	145071	770580	77.06	83.24	83.17						
908210N	Uganda	3.14	A	357155	343993	963148	96.31	100	99.81						
ULIDS-0461	Uganda	3.12	A	210084	175393	834871	83.49	100	99.47						
907856N	Uganda	2.97	A	137878	108840	789394	78.94	100	99.88						
ULIDS-0943	Uganda	2.93	A	345716	342175	989757	98.98	100	99.81						
ULIDS-0595	Uganda	2.88	A	154700	153619	993012	99.3	100	99.75						
905960N	Uganda	2.3	A	42816	31840	743647	74.36	100	99.97						
ULIDS-0955	Uganda	2.29	D/E	157565	156910	995843	99.58	84.63	84.54						
Median				**80261**	**56066**	**779909**	**77.99**	**100**	**99.75**	**3045183**	**2791553**	**956800**	**95.68**	**100**	**100**

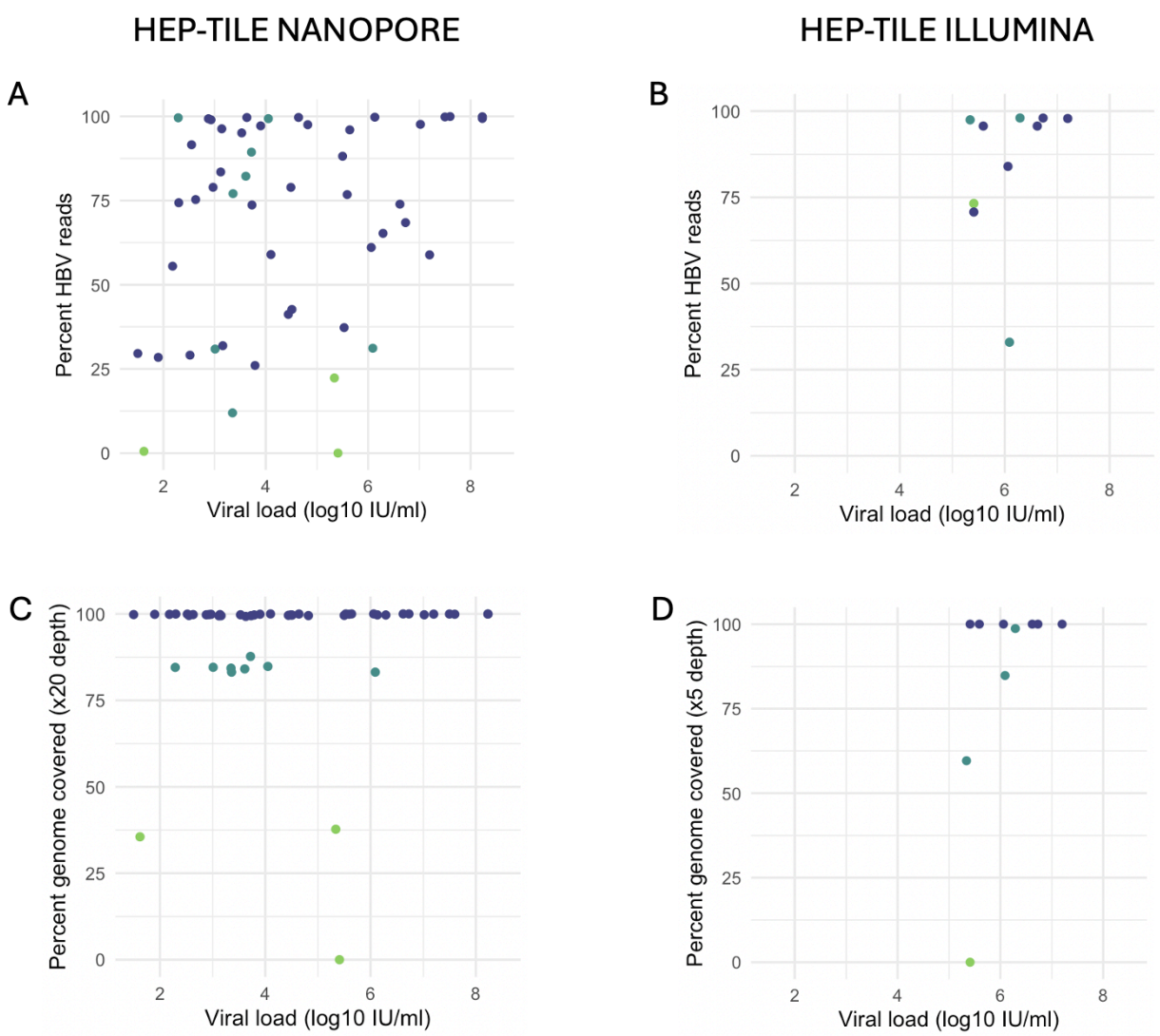


Figure 4.12. Sequencing outputs for HBV WGS using amplicon based target enrichment (HEP-TILE).

Top row: Relationship between VL and percentage of reads on-target (ie. HBV) for (A) HEP-TILE sequenced on Nanopore and (B) HEP-TILE sequenced on Illumina.

Bottom row: Relationship between VL and percentage of genome with high confidence consensus for (C) HEP-TILE sequenced on Nanopore (x20 minimum depth) and (D) HEP-TILE sequenced on Illumina (x5 minimum depth). Panels C and D show an outlier with VL 5.41 $\log_{10}$ IU/ml, with disproportionately low percentage genome coverage. Repeat VL was 3 $\log_{10}$ IU/ml.

Full genome coverage (x1 depth) was generated in 40/50 samples. Full-length HBV consensus genomes at x20 depth (the minimum required to accurately call a mutation), were generated in 40/50 samples, down to the lowest VL tested (1.5 $\log_{10}$ IU/ml). Partial genomes were obtained for the remaining 10 samples. These low coverage samples tended to have lower VL (**Figure 4.12**), although this did not reach statistical significance (p = 0.16, Mann Witney U test). I did not observe any relationship between HBV genotype and inability to produce whole genomes (**Figure 4.13**).

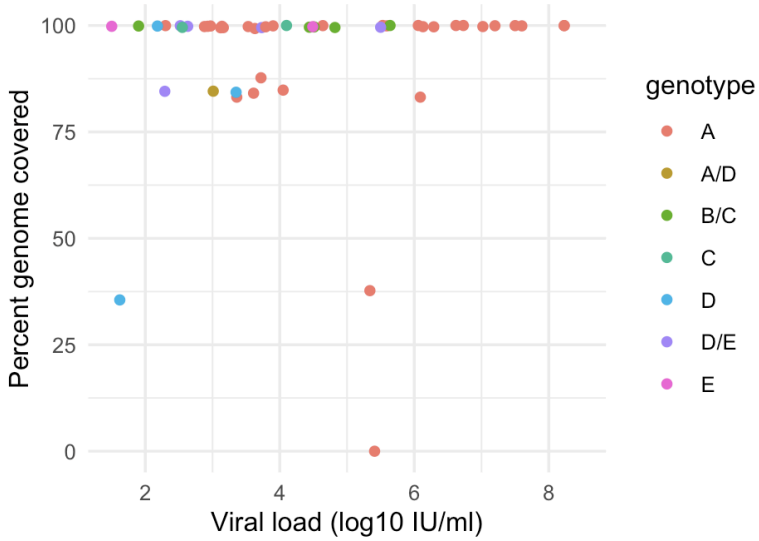


Figure 4.13. Relationship between HBV genotype and performance of HEP-TILE Nanopore workflow. Percentage genome coverage shown at x20 depth.

Sequencing clinical HBV samples enriched with HEP-TILE using Illumina

I selected a subset of 10 samples (from South Africa based on sample availability, **Figure 4.11** and **Appendix 2 section 4** for details) for HEP-TILE amplification, followed by fragmentation and sequencing on Illumina. I generated a median of 956,800 (IQR 759,122 - 977,744) HBV rpm. The relationship between HBV VL and percentage reads on target is shown in (**Figure 4.12 B and D**). Full genome coverage (x1 depth) was generated in 7/10 samples. Full-length consensus HBV genomes (minimum x5 read depth) were generated in

7/10 samples, partial genomes were obtained for 2/10 samples and the pipeline failed to generate a consensus from 1 sample due to low read coverage and anomalous mapping mismatches.

Sequencing mock HBV samples enriched with HEP-TILE using Nanopore and Illumina

I used the same 10 mock samples prepared above to validate the capture probes, in duplicate, enriching using the HEP-TILE tiled amplicon scheme and sequencing on Nanopore MinION and Illumina Novaseq X. With HEP-TILE Nanopore I generated a median of 997,016 (IQR 994,669 - 999,427) HBV rpm. With HEP-TILE Illumina I generated a median of 977,441 (IQR 963,459 - 982,631) HBV rpm sequenced (**Table 4.5**). Enrichment with HEP-TILE led to the generation of full-length HBV sequences across genotypes on both platforms (**Figure 4.14**).

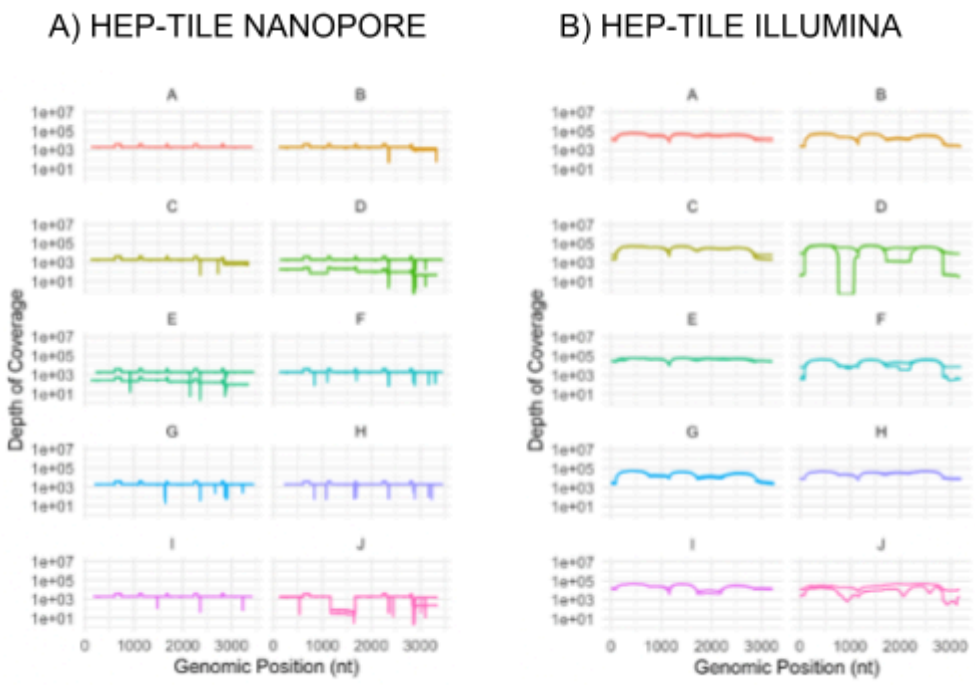


Figure 4.14. HBV genome coverage for mock samples of HBV genotypes (A-J) enriched using HEP-TILE and sequenced on (A) nanopore and (B) illumina platforms.

Table 4.5. Sequencing metrics for mock HBV samples for HEP-TILE Nanopore and HEP-TILE Illumina workflows

Genotype	Reference genome	HEP-TILE nanopore						HEP-TILE illumina					
		Total reads	Mapped reads	HBV reads per million sequenced	Percent reads mapped to HBV	Percent genome coverage (x1)	Percent genome coverage (x20)	Total reads	Mapped reads	HBV reads per million sequenced	Percent reads mapped to HBV	Percent genome coverage (x1)	Percent genome coverage (x5)
A	KU736919	95199	94946	997342	99.73	100	100	4779606	4517282	945116	94.51	100	100
		100155	100099	999441	99.94	100	100	7560102	7056345	933366	93.34	100	100
B	not published	79709	79284	994668	99.47	100	100	5013574	4923003	981935	98.19	100	100
		53460	53283	996689	99.67	100	100	4757802	4633475	973869	97.39	100	100
C	MT111596	67502	67463	999422	99.94	100	100	3920376	3773438	962519	96.25	100	100
		50962	50937	999509	99.95	100	100	4546038	4386640	964937	96.49	100	100
D	NC_003977	777	777	1000000	100.00	100	100	7559004	7386456	977173	97.72	100	100
		151565	150757	994669	99.47	100	100	4595728	4581007	996797	99.68	90	90
E	KU736895	2310	1363	590043	59.00	100	100	7866442	7537133	958137	95.81	100	100
		174563	174504	999662	99.97	100	100	7175694	6915732	963772	96.38	100	100
F	DQ823095	64275	64029	996173	99.62	100	100	4011426	3770749	940002	94.00	100	100
		68247	68081	997568	99.76	100	100	4544664	4392756	966574	96.66	100	100
G	AB056513	60886	60813	998801	99.88	100	100	4869602	4790237	983702	98.37	100	100
		78879	78761	998504	99.85	100	100	4110952	4048170	984728	98.47	100	100
H	FJ356716	83243	82928	996216	99.62	100	100	5613544	5488416	977710	97.77	100	100
		95882	95846	999625	99.96	100	100	5122012	5020249	980132	98.01	100	100
I	AB562463	255721	254408	994865	99.49	100	100	3159806	3103797	982275	98.23	100	100
		289336	284791	984292	98.43	100	100	4547926	4499098	989264	98.93	100	100
J	AB486012	83559	81982	981127	98.11	100	100	3057900	3036911	993136	99.31	100	100
		553110	549429	993345	99.33	100	100	3149202	3092929	982131	98.21	100	100
Median		**81476**	**80633**	**997016**	**99.70**	**100**	**100**	**4676765**	**4549145**	**977441**	**97.75**	**100**	**100**

4.6.4 Conclusion

The HEP-TILE sequencing protocol offers flexibility to sequence across a range of HBV VL, able to sequence full genomes down to 1.50 $\log_{10}$ IU/ml and across all known HBV genotypes. I was able to sequence whole amplicons on the ONT MinION and fragment amplicons for sequencing on Illumina Novaseq X.

The ~600-715bp amplicons used in HEP-TILE improve sensitivity compared to single amplicon schemes[131], enhancing sequencing success from low VL or degraded samples. While HEP-TILE amplicons can be sequenced on either Nanopore or Illumina, the use of the ONT native barcoding kit described here (ligation based) rather than ONT rapid barcoding kit or NEB-Ultra-II-FS kit (both requiring fragmentation) enables the sequencing of whole amplicons. Although shorter amplicon schemes (300-600bp) can facilitate transfer to Illumina platforms without amplicon fragmentation, HBV diversity and genome structure prohibit the generation of a shorter pan-genotypic scheme. Generally, shorter amplicons improve sensitivity and platform transferability, however longer 1-2 amplicon schemes or schemes generating concatenated full genomes are useful for haplotype reconstruction[143,222]. The use of two primer pools with equivolume pooling (instead of pooling individual reactions for each amplicon) simplifies the PCR step and eliminates the need for rebalancing of individual primer pools post-PCR required by other methods[223].

4.7. Methods comparison HEP-TILE v.s. Probe based capture

4.7.1 Background

Using the data generated in sections 5 and 6 above, I compared the accuracy and sensitivity of consensus genomes generated by HEP-TILE (nanopore and illumina) v.s. probe based capture (illumina only).

4.7.2 Methods

I compared the accuracy of consensus genomes from 10 mock samples (**Figure 4.15**, described in **Chapter 2**) by aligning the sequences for each genotype to the known reference sequence using Mega11 and calculating the number of mismatches, insertions, deletions and positions without a base called (N's) using R version 4.2.2[118]. I analysed a subset of 10 clinical samples from the South Africa cohort (> 5 $\log_{10}$ IU/ml) that had been sequenced using all three workflows (**Appendix 2, section 4**) so that I could compare the ability of capture vs. HEP-TILE to generate HBV WGS at VL > 5 $\log_{10}$ IU/ml.

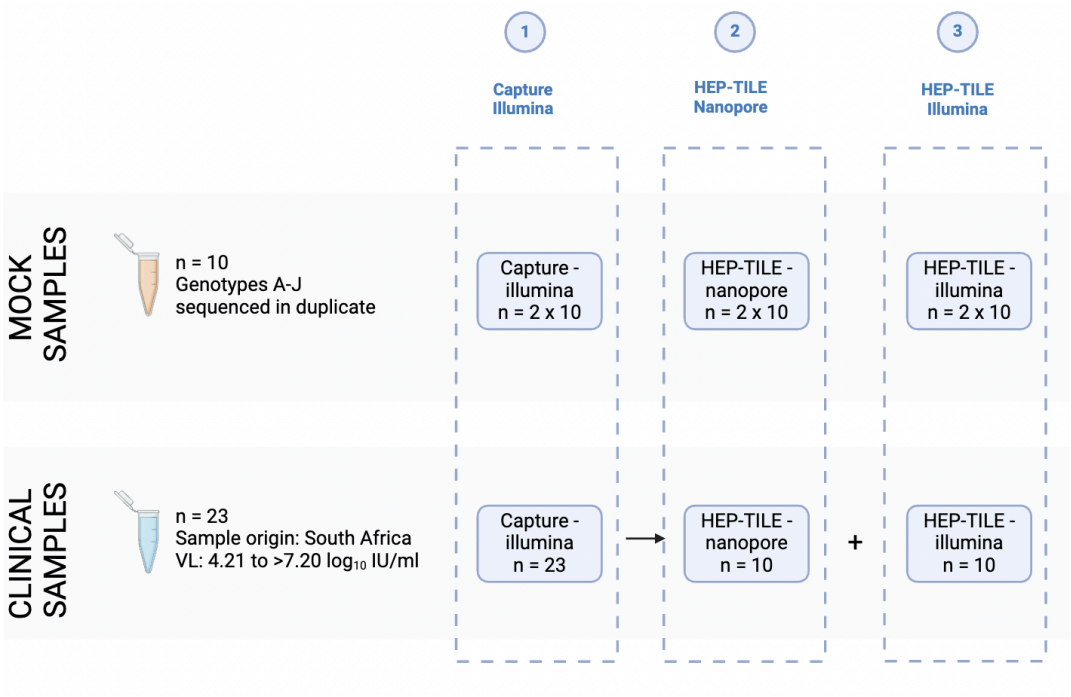


Figure 4.15. Overview of sample sets sequenced by 3 methods in sections 5 and 6, used in analysis to compare outputs of different platforms. Created in BioRender. Lumley, S. (2025) https://BioRender.com/p06t038

4.7.3 Results

Accuracy of consensus genomes

I sequenced all 10 mock samples using all three workflows, with each sample processed in duplicate (capture Illumina, HEP-TILE Illumina, HEP-TILE Nanopore). I compared consensus sequences generated from mock samples to the known plasmid sequence, demonstrating 100% accuracy of consensus genomes produced with capture-Illumina and HEP-TILE Nanopore and 99.98% with HEP-TILE Illumina. Four samples sequenced with HEP-TILE Illumina had mismatches or bases designated ‘N’, however the median number of mismatches and N’s was 0. There were no indels for any of the methods.

Comparison of capture vs. HEP-TILE ability to generate whole genomes from high viral load clinical samples

At x1 depth, capture-Illumina and HEP-TILE Nanopore generated full genomes in 8/10 samples, with HEP-TILE Illumina in 7/10 samples. At depths required to confidently call consensus (x5 for Illumina, x20 for Nanopore), full genomes were generated in 4/10 samples for capture-Illumina, 8/10 for HEP-TILE Nanopore and 7/10 for HEP-TILE Illumina (**Table 4.6**). However, the results of these 10 samples are not representative of the performance of HEP-TILE seen in **Figure 4.12**, from which I easily obtained full genomes in the majority samples >5 $\log_{10}$ IU/ml. This may reflect an issue with the VL metadata for SA samples; one of the two samples which failed to generate full sequences had residual volume available to repeat the PCR, recorded PCR was 5.41 $\log_{10}$ IU/ml whereas repeat VL was 3 $\log_{10}$ IU/ml, residual volume was not available for other poorly performing samples.

4.7.4 Conclusion

I have developed two pan-genotypic viral enrichment methods for HBV WGS: HEP-TILE and probe based capture. HEP-TILE offers the greatest flexibility to sequence across a range of HBV VL, able to sequence full genomes down to 1.50 $\log_{10}$ IU/ml (~30 IU/ml, lowest tested here).

I report percent genome coverage at x1 depth in addition to the percentage of the genome with consensus calls to allow comparison with other published methods presenting coverage at x1 depth. We use a higher threshold of x20 depth to call bases at a consensus level for Nanopore, and x5 for Illumina to improve certainty of the consensus base call (higher threshold for Nanopore used due to the higher error rate of Nanopore sequencing technology).

The results presented in sections 4.5, 4.6 and 4.7 are from an iterative set of experiments and protocol development, rather than a pre-planned head-to-head comparison. This meant I was unable to sequence all samples using all three enrichment approaches. Furthermore, the extraction method changed over time. Nevertheless, these data showcase the capabilities of each workflow.

4.8. Ultra low viral load samples

4.8.1 Background

Up to this point, I had tested the HEP-TILE protocol on samples down to VL 30 IU/ml, the lowest quantified using the commercial assays available for this cohort. Working with UKHSA and Oxford University as part of an ongoing multi-site international study, I was able to trial the HEP-TILE sequencing method on samples with VLs down to 1 IU/ml, quantified in triplicate by collaborators using an ultra-sensitive in-house qPCR[127]. This is of particular relevance to investigation of OBI, which can be associated with low VLs (<200 IU/ml)[54].

In low VL samples there is stochasticity associated with whether any given aliquot of DNA extract contains a template genome. The chance of a template being present can theoretically be improved by increasing the input:output ratio of extraction. Here I was provided with extract generated using a high volume (5ml) input[127], which I hypothesised would increase the chance of viral enrichment using the HEP-TILE tiled amplicon PCR and therefore increase the chance of generating HBV sequence at lower VL.

4.8.2 Methods

I was provided with 23 samples and one negative control (see **Appendix 2, section 5** for details) which had been extracted using the Roche large volume extraction (5ml input) and quantified using an ultra-sensitive in-house PCR[127] by the study coordinators. HBV VL was quantified by collaborators in-house at the University of Oxford[127]. I was blinded to the sample VL and the identity of the negative control. I performed the HEP-TILE PCR using two different methods (1) 10ul sample input and a 35 cycle PCR in duplicate, (2) 6ul sample input and a 40 cycle PCR. I performed library preparation and sequencing according to the protocol I have developed and optimised as described above. After submitting raw reads,

consensus genomes and sequencing methods to the study coordinator, I was then un-blinded to the sample VL and was able to analyse the relationship between VL and percentage genome coverage at x20 depth.

4.8.3 Results

I was able to sequence HBV in samples with VL as low as 1 IU/ml. Protocol 1 (10ul input, 35 cycle PCR) identified HBV in 13/23 samples (5 full genome, 8 partial genome, protocol 2 (6ul input and 40 cycle PCR) identified HBV in 18/23 samples (4 full genome, 14 partial genome). The negative control had 0% HBV genome coverage at x1 depth. The correlation of VL with percent genome coverage at x20 depth is shown in **Figure 4.16 A**, and a direct comparison of the percentage genome coverage at x20 depth between the two sequencing protocols is shown in **Figure 4.16 B**.

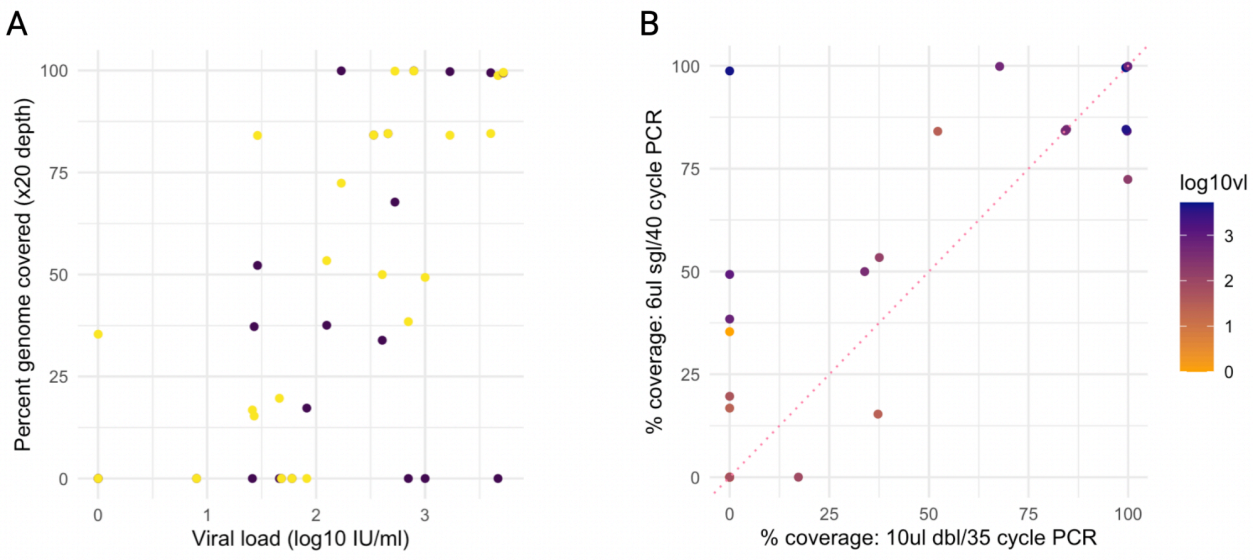


Figure 4.16. Sequencing metrics from low VL samples extracted using high volume extraction. A) Percent genome coverage by VL, purple dots = 10ul duplicate reaction with 35 cycle PCR, yellow = 6ul single reaction with 40 cycle PCR. B) Comparison of percent genome coverage at x20 depth between the two PCR methods, pink dotted line shows line of equivalence between the two PCR protocols.

4.8.4 Conclusion

I was able to generate full genomes from samples with VL down to ~2 $\log_{10}$ IU/ml and partial genomes down to 1 IU/ml which I had not previously been able to test as we were using commercial HBV VL assays with a lower limit of quantification of 20-30 IU/ml. Method comparisons are hampered due to insufficient extract volume to perform technical replicates, and the stochasticity associated with the generation of sequencing data at very low VL. The 40 cycle PCR (protocol 2) more consistently generated partial/full genomes. Counterintuitively, more samples failed PCR when I used a high volume (10ul) sample input (protocol 1), I had not seen this previously, but a hypothesis would be that a component of the high volume extraction kit was inhibiting the PCR reaction. The high volume extraction did not substantially improve percentage genome coverage of samples with VL 1.5 - 3 $\log_{10}$ IU/ml previously tested by other methods (**Figure 4.12),** caveated that I was unable to perform a direct comparison of extraction methods as I received the DNA extracts rather than original samples.

4.9. HBV WGS from dried blood spots

4.9.1 Background

The costs and logistics surrounding storage and transport of plasma or serum samples is challenging when working in remote settings. The use of alternative sample types such as dried blood spots (DBS) is appealing due to the ease of obtaining finger-prick specimens, lack of cold chain requirement and simplified transport and storage. DBS has shown to be a feasible sample type for performing HBV serology and HBV VL[224–227]. Successful sequencing of single sub-genome HBV amplicons has been demonstrated previously from DBS, but there are no reports of HBV WGS[225,228,229].

4.9.2 Methods

I collaborated with a group working in the Democratic Republic of Congo[125] (led by US team Prof Peyton Thompson and Prof Jonathan Parr, with PhD student Camille Morgan) to sequence a set of HBV samples including 8 DBS samples (see **Appendix 2 section 6.1** for details), they had collected as part of the Horizontal and Vertical Transmission of Hepatitis B (HOVER-HBV) study (see **Methods** for cohort description). I trialled the primer scheme I was using at the time, hbv/500/v1.1.0[139] with the ONT library prep described in the HEP-TILE protocol[149] to generate WGS.

4.9.3 Results

I successfully obtained full genome sequences from 6 out of the 8 samples sequenced (VL 3.78 - 6.08 log_{10} IUml) (**Figure 4.17**; DBS samples outlined in orange) with partial genomes in 2/8 samples (lower viral loads; 3.56 and 3.53 log_{10} IUml) related to the ‘amplicon 4’ dropout

in the initial hbv/500/v1.1.0 primer scheme described in HEP-TILE methods above. Full results for the cohort are described in **Chapter 6.**



Figure 4.17 HBV genome coverage using a tiled amplicon protocol from the HOVER-HBV study in Democratic Republic of Congo. DBS samples are highlighted in orange. Graphs show individual amplicons on the x axis and depth of coverage on the y axis, amplicons with above x20 coverage are shown in teal and less than 20x depth are highlighted in red.

4.9.4 Conclusion

DBS is a feasible alternative substrate for HBV WGS, this could be useful prospectively when collecting and storing samples for monitoring of viral evolution, drug resistance and vaccine escape mutants but also retrospectively to look at archived sample sets to study the phylogeography of HBV in settings where this has not previously been feasible. In the future I could trial the current iteration of the primer scheme hbv/600/v2.1.0, which solves the issue of the amplicon 4 dropout, to demonstrate the generation of full genome sequences in lower VL samples.

4.10. Discussion

I have developed three viral enrichment methods for HBV WGS: HEP-TILE, probe based capture and gPCR-RCA. Although both capture and HEP-TILE are able to effectively enrich HBV genomes from all known genotypes (A-J), as evidenced by application to mock samples, HEP-TILE offers flexibility to sequence across a range of HBV VL, able to sequence full genomes down to 1.5 $\log_{10}$ IU/ml and partial genomes down to 1 IU/ml. Although our capture approach can be used to improve the efficiency of viral sequencing at VL >~5.50 $\log_{10}$ IU/ml, its efficacy for enriching HBV genomes and producing full-length consensus genomes diminishes at lower VLs, as observed elsewhere[220]. RCA was only feasible at VL > 6 $\log_{10}$ IU/ml, with the gPCR-RCA adaption required to sequence samples of VL 5 $\log_{10}$ IU/ml.

4.10.1 Real world application of methods

Table 4.7 compares the pros, cons and costs of each workflow, estimated per sample based on an initial batch of 1000 samples. Cost per sample decreases with multiplexing (e.g. 96 samples for Nanopore, more possible for Illumina). Due to the initial cost of purchasing capture probes (~£6300 for this panel), sequencing smaller sample sets with capture becomes prohibitively expensive. Costs for Illumina sequencing vary significantly depending on whether the option of purchasing a partial lane on a high throughput platform is available.

The bioinformatic pipelines used here generate consensus sequences and therefore do not identify minority variants (quasispecies), this is an area for future work. Further studies are required to determine the clinical relevance of minority drug resistance variants in affecting treatment outcomes for HBV.

Although full genome HBV sequencing coverage should be achievable using HEP-TILE, if samples at extreme ends of the VL spectrum are pooled on the same run, uneven read distribution is expected between samples, with loss of coverage for the lowest VL samples. Therefore batching samples by VL and normalising the mass of PCR product being taken forwards into library preparation is recommended.

Table 4.7. Summary of four HBV WGS pipelines, comparing time, cost and outputs.

	gPCR-RCA - Nanopore	Capture - Illumina	HEP-TILE - Nanopore	HEP-TILE - Illumina
Time required for enrichment and library prep	2 days (RCA overnight)	1.5 days (capture overnight)	1.5 days (PCR overnight)	1.5 days (PCR overnight)
Sensitivity (HBV DNA VL) for generating HBV WGS	5 $\log_{10}$ IU/ml lowest tested	> 5.5 $\log_{10}$ IU/ml	> 1.5 $\log_{10}$ IU/ml	> 1.5 $\log_{10}$ IU/ml
Consensus sequence accuracy*	99.7% using previous flow cell chemistry, likely higher.	100%	100%	99.98%
Cost per sample (enrichment & library prep reagents)**	£19	£29	£11	£19
Cost per sample* (sequencing)**	£5.60 E.g. 96 samples multiplexed per flow cell, bulk purchasing 12 flow cells, 100,000 reads per sample	£25 E.g.multiplexed on partial lane on Novaseq X, 1 million reads per sample	£5.60 E.g. 96 samples multiplexed per flow cell, bulk purchasing 12 flow cells, 100,000 reads per sample	£2.50 E.g. multiplexed on partial lane on Novaseq X, 100,000 reads per sample
Key attributes	Full length viral haplotypes, portability	Fewer PCR cycles (13+12) than HEP-TILE (35 - 40)	Sensitivity at low VL, portability, cost	Sensitivity at low VL
Key concerns	Time Sensitivity at low VL, PCR may bias quasispecies	Cost prohibitive in smaller batches, cannot reconstruct full genome haplotype	Cannot reconstruct full genome haplotype, PCR may bias quasispecies	Optimisation of fragmentation time to PCR product concentration required, cannot reconstruct full genome haplotype, PCR may bias quasispecies

All costs are based on cost of reagents in the UK in 2024 (GBP) and do not include staff or equipment costs. VL - viral load; gPCR - full length genomic PCR; RCA - rolling circle amplification; PCR - polymerase chain reaction.
*based on sequencing of mock samples with known sequence
** reagent costs for enrichment and library prep reagents calculated in GBP for the first 1000 samples. Cost excludes extraction and QC (Qubit/Bioanalyser reagents).
*** NB price extremely variable - depends on platform used and whether performing in-house or outsourcing commercially. Excludes hardware (i.e. Laptop/MinION for Nanopore, and Novaseq/Miseq for Illumina)

Viral enrichment by PCR is highly sensitive to amplicon contamination therefore caution should be taken to keep work areas, equipment and reagents free of contamination, and multiple negative controls should be included. To avoid *in-silico* 'contamination' double barcoding is required when demultiplexing to eliminate chimeric reads which may otherwise lead to mis-assignment of reads to incorrect barcodes[230]. These wet lab and bioinformatic principles are also applicable to sequencing other targets.

In theory, use of a high volume extraction should increase the sensitivity of HBV sequencing, although I did not observe a large impact. However it should be noted that in the majority of clinical or research settings, high volume (5ml) extraction will not be feasible as usually < 500ul serum or plasma is available for sequencing.

4.10.2 Caveats and limitations

The probes and primers were designed to be pan-genotypic, however only take into account the diversity of sequences on HBVdb, which contains relatively few genotype F-H sequences, and under-represents certain global populations[32]. A strength of my work is the use of mock samples, which allowed me to assess performance across all genotypes, including rarer genotypes F-J that were unlikely to be present in our cohorts, however might still be present in other settings globally. Analysis of rarer genotypes is often missing in reports of sequencing schemes, however will become increasingly important as more HBV WGS is performed from diverse global regions, and known HBV sequence diversity expands. While I report these methods as pan-genotypic using mock samples generated from plasmids containing HBV genotypes A-J, it is important to note that I only tested one variant per genotype. Some genotypes have multiple subtypes; future studies could test enrichment methods with a wider range of subgenotypes.

Sequencing low VL samples < 4log_{10} IU/ml has inherent stochasticity, depending on the presence of viral DNA in a sample or extract, and assay chemistry. Future work could test the impact of pooling products from PCRs performed in duplicate or triplicate and compare the ability of capture or HEP-TILE to accurately quantify low abundance minority variants, requiring a panel of control material at varying VLs with known variant percentages, sequenced with multiple technical replicates. Finally, enrichment methods could be trialled on different sample types, for example liver biopsies where high host background is a challenge.

4.10.3 Challenges

Optimising methods sequencing was challenging because of the limited number of samples and restricted sample volume available. This influenced the number of comparisons between methods that could be performed and the number of achievable replicates of each protocol which would have been useful to assess reproducibility of methods and ensure robust comparisons between enrichment/library prep optimisations. I considered using mock samples (plasmids diluted in non-infected serum) or high VL samples serially diluted in non-infected serum to perform the optimisation work, but given the hypothesised impact of HBV rcDNA structure on target enrichment success, and the difference in HBV genome structures found in individuals with high v.s. low VL, I decided against.

Correlation between VL and percentage genome coverage was more variable than anticipated. Factors such as storage duration, shipping conditions, and repeated freeze-thaw cycles may all influence sample quality, although DNA is generally robust. Poor sample quality or inaccurate VL reporting (different sample/timepoint tested or lab/reporting error) may explain instances where samples that were initially reported to have high VL yielded lower coverage than expected (e.g. sample 1116 for which VL was reported as 5.41 log_{10} IU/ml but on repeat on our sample was 3 log_{10} IU/ml).

Finally, although I demonstrate that it is feasible to sequence fragmented amplicons on Illumina platforms, the main challenge taking this forwards will be tailoring the fragmentation time to the concentration of PCR product to avoid over-fragmentation of low concentration samples.

4.10.4 Alternative methods

At the same time as I was developing HEP-TILE, two other groups developed tiled amplicon methods for HBV nanopore sequencing[222,231]. Ringlander et al[222] developed a two amplicon scheme with primers designed in-house, and Tshiabuila et al[232] developed a ten amplicon scheme using the original PrimalScheme primer designer[233]. Neither of these schemes are pan-genotypic, designed only for genotypes A-E, however the Ringlander et al protocol has good sensitivity and the long amplicons provide an attractive option of using a nested PCR to generate a partial genome sequencing of a Pol/S fragment if the whole genome sequencing is unsuccessful. The Ringlander et al protocol amplifies both HBV DNA and RNA using Superscript III polymerase, which may contribute to the increased sensitivity, in addition to the nested PCR. I removed the RT step from the adapted SARS-CoV-2 protocol to ensure that sequencing results were representative of serum DNA and not confused by combined DNA-RNA sequencing. The Tshiabuila et al[231] protocol generates HBV sequences with a gap, as it does not include an amplicon to span position ‘0’, and has the same issue of amplicon drop out at the negative strand nick, which I observed and addressed by adjusting my primer position.

Both of these protocols use the ONT rapid barcoding kit for library preparation, a tagmentation based method where a transposase randomly cuts amplicons and adds a barcode. I used the ONT native barcoding kit in which a ligase adds a barcode to the end of the amplicon, enabling full length genomes to be sequenced. The major advantages of rapid

barcoding are the speed (10 mins v.s. 2.5 hours for native barcoding) and lower cost. The disadvantages include the generation of shorter reads, reported lower read count, and inability to perform stringent demultiplexing with dual barcoding since barcodes are only added to one end. Furthermore there is a risk of barcode cross-over between samples due to residual transposase activity after pooling, which is harder to identify as only 1 barcode is present per DNA strand.

4.10.5 Conclusion

Having a toolbox of diverse sequencing methods allows the most appropriate method to be selected according to local infrastructure, cost and throughput requirements (**Table 4.7**). Collectively, these protocols will facilitate the wider generation of HBV sequence data, although HEP-TILE in particular is readily transferable to other settings, capitalising on the rapid global expansion of technical expertise and infrastructure for amplicon based WGS that occurred during the SARS-CoV-2 pandemic. This data will lead to a more accurate and representative global picture of HBV molecular epidemiology, cast light on persistence and pathogenesis, and enhance understanding of treatment responses and clinical outcomes. These aspirations align with high profile international goals for elimination of HBV as a public threat.

Chapter 5: HBV drug resistance

5.1. Availability of content

This chapter combines outputs from three manuscripts, I am the first author of all 3 papers and wrote the original draft of all the manuscripts:

1. A case report which summarises reasons for lack of virologic suppression in HBV (published in BMJ sexually transmitted infections[234]);
2. A systematic review and meta analysis of drug resistance in HBV infection, published in Journal of Clinical Virology[100]
3. An opinion/perspective paper that is accepted at Lancet Microbe. In this, I summarise outputs from a multidisciplinary consultation convened by the WHO in 2023 to discuss HBV drug resistance (HBVDR);

5.2. Background

NAs suppress HBV DNA to below quantifiable thresholds in the majority of people receiving treatment. However, viraemia persists in a proportion of those offered treatment (up to 20% after one year in a recent population analysis[235]). Inadequate virological response to therapy may be attributed to a range of influences including host, viral and environmental factors, some of which may interact and influence the selection of RAMs (**Table 5.1)**. HBV sequencing is the gold standard agnostic method for detecting known and novel RAMs. Resistance is predictably selected by exposure to 3TC, which also influences susceptibility to ETV and ADV[44]. The emergence of resistance to TFV is thought to be infrequent[57,74]. The extent to which TFV resistance is a real-world problem - either for individual patients or on public health grounds - remains uncertain and is likely to vary between population settings according to risk factors including prior NA exposure.

Although the risk of HBVDR as a threat to treatment outcomes at a public health level is currently thought to be low, there is a potential for HBVDR to influence individual outcomes, particularly among those who are treatment-experienced. As therapy reaches more diverse populations it is important to establish a baseline of HBVDR from existing data, and to ensure that appropriate prevention, monitoring and surveillance approaches for HBVDR are in place as test-and-treat interventions are scaled up, to coincide with investment in the global scale-up of HBV diagnosis and treatment.

In this chapter I begin by describing a case which highlights the complex interplay of factors that influence outcomes of drug therapy. I then focus on HBVDR as one of the factors that influences the outcome of therapy, performing a systematic review and meta-analysis of the risk of ETV and TFV resistance over time in those with/without prior NA exposure. This is the first meta-analysis to report on ETV and TFV resistance. I finish by reporting outputs from a multidisciplinary consultation convened by the WHO in 2023 to discuss HBVDR. I consider the HBVDR evidence gaps, and propose a set of recommendations for closing the current research gaps to inform next steps for the field.

Table 5.1: Factors that potentially underlie inadequate virological response to therapy for hepatitis B virus (HBV) with nucleos(t)ide analogue (NA) agents

Domain	Factors influencing HBV treatment outcomes
Viral biology	> High HBV VL at treatment outset > Poor drug/target avidity, either due to specific RAMs or to other polymorphisms that influence susceptibility > Drug resistance due to the selection of specific RAMs
Host biology and pharmacokinetics	> Age, presence of liver disease, body mass index > Genetic determinants of drug handling > Malabsorption > Alterations in drug distribution or tissue penetration > Drug-drug interactions / polypharmacy
Psychosocial and economic factors	> Out-of-pocket costs for drugs, laboratory testing, therapeutic monitoring, clinic visits > Vulnerable and marginalised characteristics of 'key populations' > Stigma and discrimination presenting barriers to consistent therapy > Insufficient duration of treatment > Limited duration of treatment available through publicly funded programmes in some settings
HBV programme factors	> Interruptions to treatment due to drug stock-outs > Poor programme retention > Inappropriate NA treatment prescribed > Inadequate access to HBV VL monitoring > Centralised care, causing logistical barriers to access > Insufficient training in HBV management for healthcare personnel (e.g. poor application of HBV VL test results to inform management) > Low education and advocacy for HBV

NA – nucleos(t)ide analogue; RAM – resistance associated mutation; VL – viral load

5.3. Aims

Aim 5.1. To illustrate the interplay of factors that influence outcomes of HBV nucleos(t)ide analogue therapy with a case report

Aim 5.2. To perform a systematic review and meta-analysis to summarise the risk of drug resistance to ETV and TFV in people receiving treatment for CHB, determining key differences in patterns of resistance between agents and prior nucleos(t)ide analogue exposure

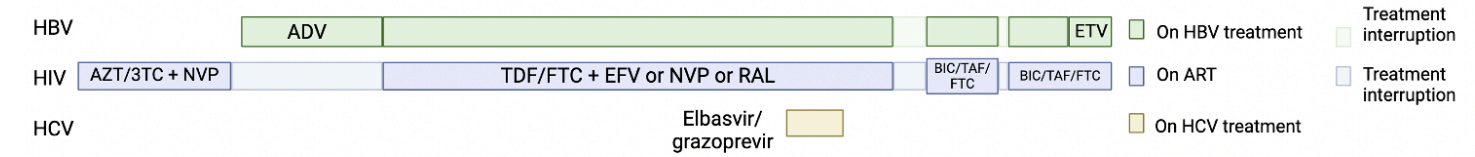
Aim 5.3. To summarise key concepts and evidence gaps relating to HBVDR to inform next steps for the field

5.4. Case report - factors influencing outcomes of NA therapy

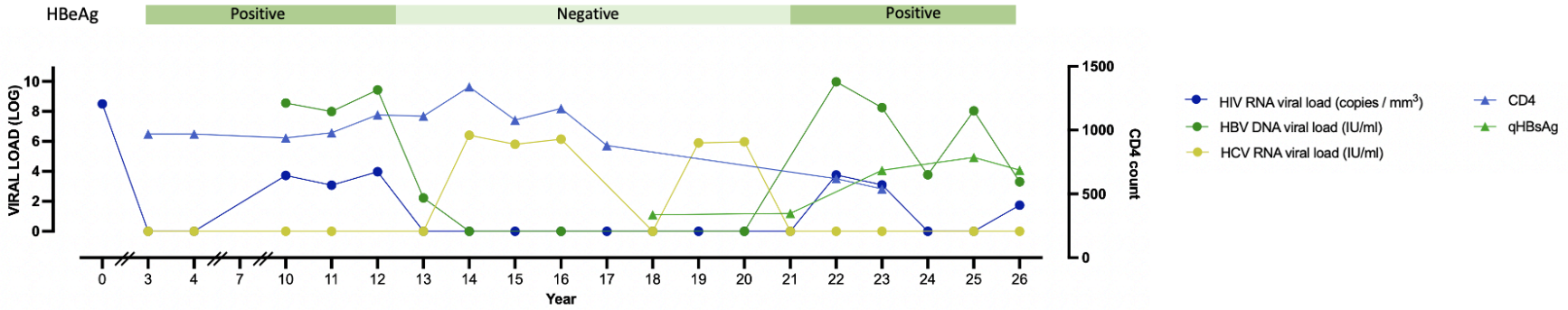
I reported on an adult male receiving treatment for HBV and HIV co-infection over a period of 26 years in a central London clinic (**Figure 5.1**)[234]. The clinical team obtained written informed consent; additional samples and data were collected with ethical approval (ref. 11/LO/0421). I retrospectively reviewed data from his routine clinical records and performed HBV WGS.

At initial diagnosis HBV VL was 8.5 $\log_{10}$ IU/ml and HBeAg was positive. He screened negative for HDV infection. Antiretroviral therapy (ART) was commenced, with HIV suppression from 3.7 $\log_{10}$ RNA copies/ml to undetectable. Subsequently HBV was diagnosed and HBV VL progressively suppressed on HBV-active regimens containing 3TC (from baseline), switched to ADV (year 8), and then TFV (year 12), prescribed in line with changing guidelines (**Figure 5.1 A and B**). HBeAg was undetectable by year 13 (although he did not develop anti-HBe). He received successful intercurrent treatment for HCV infection.

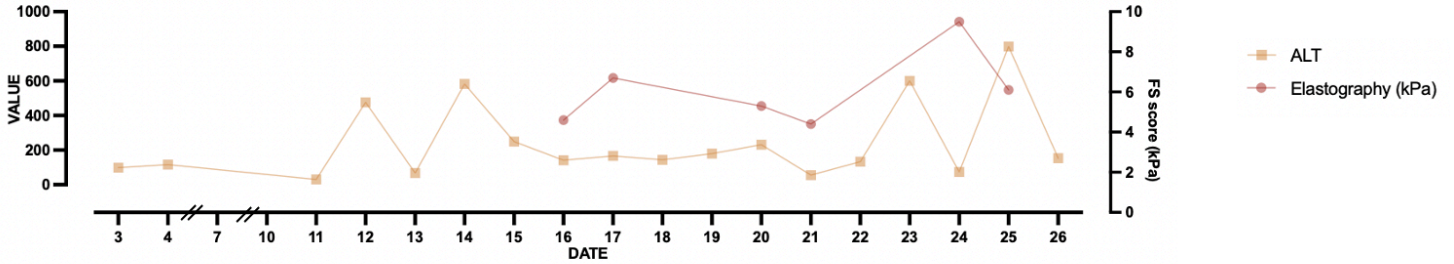
A. Antiviral treatment history



B. Viral markers



C. Liver biochemistry and elastography



D. HBV sequencing

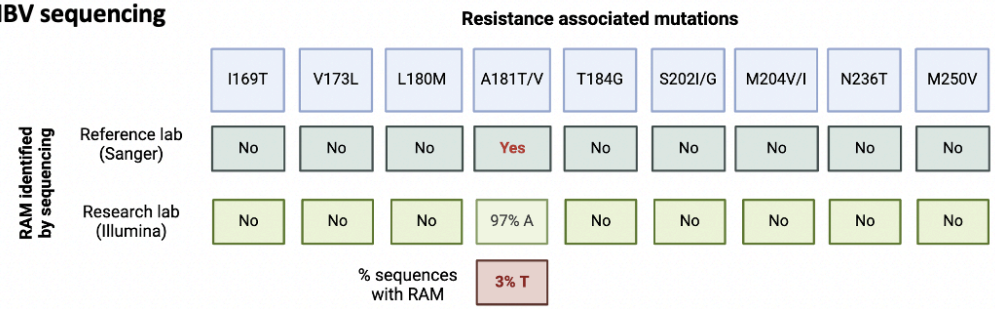


Figure 5.1. Summary of treatment and laboratory results from a case report of an individual with non-suppression of HBV VL despite being prescribed antiviral therapy.

A) Timeline showing antiviral treatment for HBV, HIV and HCV, showing years since diagnosis on x axis. B) Timeline showing trends in VLs for HBV, HIV and HCV, in addition to CD4+ count, quantitative HBsAg and HBeAg status (NB x-axis not to scale). C) Timeline showing ALT and liver elastography scores (NB x-axis not to scale). D) Resistance associated mutations (RAMs) listed in EASL guidance[57] mapped to the HBV polymorphisms identified in HBV sequenced from this individual through Sanger and Illumina sequencing. ADV - adefovir; ETV - entecavir; AZT - zidovudine; 3TC - lamivudine; NVP - nevirapine, TDF - tenofovir disoproxil, RAL - raltegravir, TAF - tenofovir alafenamide; BIC - bictegravir; FTC - emtricitabine.

Due to illness and emergency hospital admissions unrelated to blood borne virus infection, he was unable to sustain NA treatment, with documented gaps at intervals between years 22 and 25. HIV and HBV rebounded, with associated liver inflammation (peak ALT 902 IU/L), and HBeAg status reverted to positive. On reinstating HBV-active ART (including TFV and emtricitabine), HIV VL suppressed within 1 month, but HBV VL remained persistently elevated between 3.3-9.9 $\log_{10}$ IU/ml (**Figure 5.1 B and C**).

A clinical diagnostic laboratory identified the presence of a mixed genotype A and G infection with a resistance mutation at position 181 in the HBV RT sequence (A181T) and reported potential drug resistance to 3TC, ADV and telbivudine (**Table 5.2**). Analysis of full genomes generated by HBV probe based capture and Illumina sequencing (methods described in **Chapter 4**) demonstrated dual infection with HBV genotypes A and G, and confirmed the A181T polymorphism, although only in a minority of quasispecies (**Figure 5.1 D, Table 5.3**). Alongside support to optimise adherence, ETV was added (in keeping with clinical guidelines[23657]), followed by an HBV virologic response to 2.4 $\log_{10}$ IU/ml after 1 month and 3.2 $\log_{10}$ IU/ml at 4 months.

Table 5.2. HBV drug resistance report from a clinical diagnostic virology laboratory, based on a serum sample submitted during VL rebound (year 25).

Polymorphisms in HBV reverse transcriptase (RT) codons 91-285 are listed, and summary of the interpretation of the resistance profile as reported by clinical diagnostic virology laboratory (based on https://hbv.geno2pheno.org/).

Genotype	Polymorphisms in RT domain	Clinical interpretation of resistance associated mutations
Dual infection with genotypes A and G	I53IS, T54A, I103V, N118D, I121S, N122H/I/L, N123D, N124H, M129L, W153R, V163I, A181T, V207I/M, L217R, I253V, V266A/I/T	• Lamivudine – resistant (181T) • Telbivudine - resistant (181T) • Adefovir – resistant (181T) • Entecavir – susceptible • Tenofovir - susceptible

A number of factors can collectively reduce the success of NA treatment for HBV, including high baseline VL, HIV coinfection (through exposure to 3TC (and other older NA agents), HIV-mediated immunosuppression and overlapping psychosocial disadvantage), exposure to historic regimens with low genetic barrier to resistance, therapy interruptions, physical and mental health comorbidities and complex barriers to continuity of care. These are interrelated and are likely to have a cumulative influence. Individuals with these characteristics may be in vulnerable or marginalised groups, are unlikely to be eligible for clinical studies, and may experience social stigma and discrimination. These influences mitigate against their inclusion in laboratory data, clinical cohorts and trials[237].

Table 5.3. Resistance associated mutations (RAMs) listed in EASL guidance[57], Genotype A/G consensus and polymorphisms identified in HBV isolated from this individual.

RAM	Reference sequence (NB. same aa for Geno A/G at all given sites)	Patient HBV sequence (UCLH report)	Patient HBV sequence (Illumina data)		Clinical interpretation from EASL guidance (if detected)			
			Consensus	Minority variant	3TC	ETV	ADV	TDF/ TAF*
I169T (in combination)	I		I (100%)					
V173L (in combination)	V		V (99%)	G (1%)				
L180M	L		L (100%)					
A181T/V	A	A, T	A (97%)	T (3%)	I	S	R	I
T184G (in combination)	T		T (100%)					
S202I/G (in combination)	S		S (100%)					
M204V/I	M		M (100%)					
N236T	N		N (100%)					
M250V (in combination)	M		M (100%)					

The amino acid substitution profiles are shown in the left column and the level of susceptibility is given for each drug: S (sensitive), I (intermediate/reduced susceptibility), R (resistant).
ETV, entecavir; TDF tenofovir disoproxil fumarate; TAF, tenofovir alafenamide; 3TC, lamivudine; ADV, adefovir
*In vitro data for tenofovir, in vivo data for TDF, no clinical data for TAF.

The contribution of drug resistance is doubtful in our patient, as A181T alone is not a recognised cause of TDF resistance[96]; a resistant phenotype would typically require multiple associated RAMs. More clinical and *in vitro* data are still needed to ascertain the relative contributions of different RAMs, alone and in combination including the threshold at which minority variants are clinically relevant. Adding HBV active agents, and providing adherence

support, resulted in progressive reduction in viraemia over time to <2000 IU/ml, but not to undetectable levels. Lower adherence has been associated with rising HBV and VBT in CHB [238]. In this individual, periods of non-adherence led to a more rapid re-bound of HBV DNA compared to HIV RNA, and a slower suppression when treatment resumes, suggesting that higher drug levels may be required for HBV suppression compared to HIV.

Persistent HBV viraemia poses risks for transmission, and long-term inflammatory/fibrotic liver disease (indicated by elevated ALT in this case), highlighting the need for intervention. Service improvements should focus on flexible, patient-centric access to information, consistent supplies of medication, avoiding out-of-pocket costs, and providing access to peer support, particularly for those coping with other health and/or social challenges and for whom there are barriers to care access. As HBV treatment eligibility expands and we work towards elimination targets, research is needed to better determine the factors that contribute to the presence and impact of persistent viraemia, and to optimise surveillance and clinical intervention, which I discuss further below.

5.5. A systematic review and meta-analysis of the risk of hepatitis B virus (HBV) resistance in people treated with entecavir or tenofovir

5.5.1 Background

The primary objective of this systematic review and meta-analysis is to summarise the risk of drug resistance to ETV and TFV in people receiving treatment for CHB, determining key differences in patterns of resistance between agents and prior NA exposure. This is the first and only meta-analysis to report on ETV and TFV resistance.

5.5.2 Methods

PRISMA checklist and PROSPERO registration

I set out to summarise the risk of HBV genotypic drug resistance in people receiving treatment for CHB exposed to ETV or TFV for at least 48 weeks. I conducted a systematic review and meta-analysis using the Preferred Reporting Items for Systematic Review and Meta-analysis (PRISMA) guidelines; a PRISMA checklist is provided in **Appendix 3: Supplementary Table 9**. The review protocol can be found under PROSPERO registration number CRD42023424125. Members of the Matthews group helped to screen abstracts, full papers and extract data (see manuscript for author contributions[100]), I performed data analysis and wrote the manuscript.

Data sources and search strategy

The following bibliographic databases and trial registries were initially searched by an information specialist (Eli Harriss, Bodleian libraries) on 01/06/2023 and updated 29/08/23 for studies published from database inception to the search date: Ovid MEDLINE, Ovid Embase, Ovid Global Health, Scopus, the Cochrane Central Register of Controlled Trials,

clinicaltrials.gov, the ISRCTN Registry, and the WHO International Clinical Trials Registry. Eli Harriss searched the databases using relevant index terms and free text terms, synonyms, and phrases in the title and abstract fields for relevant papers to meet the aims of this review. The full strategies are detailed in **Appendix 3: Supplementary Text 1**.

Selection criteria

Inclusion criteria were:

(i) study of ≥10 people living with CHB exposed to TFV and/or ETV monotherapy for ≥48 weeks, *and*

(ii) study reports risk of clinical resistance for either TFV and/or ETV, *and*

(iii) study reports genotypic resistance for either TFV and/or ETV, based on viral sequencing.

Exclusion criteria were:

(i) not an original research article,

(ii) incorrect study population (e.g. condition not CHB, individuals not taking TFV or ETV monotherapy, all individuals non-suppressed on TFV/ETV treatment, clinical phenotype not reported, unable to separate naive from experienced outcomes)

(iii) sample size <10 individuals (within the synthesis group of interest (ie. NA naive vs. experienced, TFV vs. ETV)),

(iv) treatment duration <48 weeks or not specified,

(v) viral sequencing not used or HBV not sequenced after TFV/ETV treatment,

(vi) not in English.

Exclusion criterion (ii) aimed to exclude studies that could potentially introduce selection bias into our estimate of risk, for example cohorts of individuals who all have VBT on TFV/ETV, or cohorts where all individuals have been referred for resistance testing. We excluded

non-English papers as they were few in number (n=5) and originated from geographical regions (Western Pacific Region) already well-represented in the data.

Screening and data extraction

The full screening strategy and data fields are described in the published manuscript[237]. The key outcome measure sought was the number of individuals developing genotypic resistance. We also collected data on the number of individuals with clinical resistance and the number successfully sequenced (**Figure 5.2**). All included studies looked for genotypic resistance in individuals with VBT. Some studies looked for genotypic resistance in all viraemic individuals (ie. individuals with VBT, primary non-response and partial response). The impact of sequencing criteria on risk estimates was explored in the sensitivity analysis.

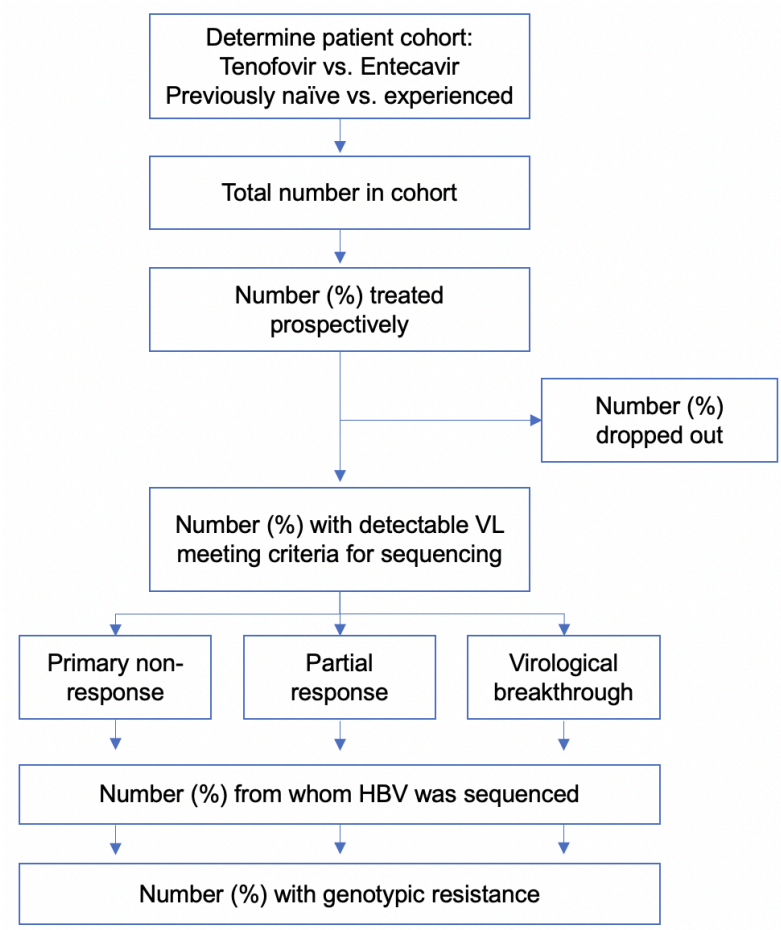


Figure 5.2. Gold standard data flow to determine risk of clinical and genotypic resistance.

Where resistance risk was reported at multiple time points within one paper, the latest time point was collected. If genotypic resistance risk was not reported for naive and experienced individuals separately for the entire trial duration, data from the latest time point at which naive/experienced individuals were reported separately was used. Where multiple doses of NA were compared, data for the cohort receiving the current licensed dose was extracted (typically 300mg/day of TDF, 25mg TAF, 0.5mg ETV (if NA naive) or 1mg ETV (in the setting of 3TC resistance). Where outcomes were not described for each dose regimen separately, the pooled results were extracted and regimen noted. Where TAF and TDF were being compared, risk of resistance was pooled. Where study details were not available in the full

text or supplementary materials, we reviewed previously published reports of the same study cohort to complete the metadata.

Risk of bias was assessed independently by two reviewers using a modified five question Joanna Briggs Institute quality assessment tool. An answer of “Yes” to all 5 questions equated to a high quality study with low risk of bias. Papers scoring 4 or 3 “Yes” answers ranked as moderate quality/moderate risk of bias and 2 “Yes” as low quality/high risk of bias. Papers scoring only 1 “Yes” were considered very low quality/ high risk of bias and were excluded. Risk of publication bias was alleviated by a comprehensive search strategy. Publication bias was assessed with funnel plots of study size against log odds (= ln(cases +0.5 / (sample size +0.5) - (cases+0.5)))[239], with the addition of 0.5 to cases and sample size to allow studies with zero events to be plotted.

I analysed our data in four groups determined a priori. In order to decide which studies were eligible for each synthesis, I compared the study population (naive vs. experienced) and NA (ETV vs. TFV) against the planned groups for each synthesis.

Definitions of drug resistance

I used standardised nomenclature for clinical, phenotypic and genotypic resistance to NA therapy [78]**.**

‘Clinical resistance’ is defined in individuals who are adherent to consistent therapy with:

(i) Virologic breakthrough (VBT) (an increase in serum HBV DNA by ≥1 log10 IU/mL above nadir on ≥2 occasions 1 month apart);

(ii) Primary non-response (inability of NA treatment to reduce serum HBV DNA by ≥1 log10 IU/mL after the first 6 months of treatment); or

(iii) Partial response (detectable HBV DNA during continuous therapy).

'Phenotypic resistance' is defined as the 'decreased susceptibility of an HBV polymerase to an antiviral treatment in vitro'[78].

'Genotypic resistance' is defined as viral populations bearing amino acid substitutions in RT that have been shown to confer resistance to antiviral drugs in phenotypic assays. These mutations are usually detected in individuals with VBT but may also be present in those with persistent viremia[78] (**Table 2.7 (Chapter 2)**). Sequencing of the PCR amplified product is the accepted standard for characterising RT mutations[78].

In my original protocol I planned to validate the presence of drug resistance against a pre-specified list of RT mutations, evidenced through phenotypic assays, using a standardised numbering system (**Table 2.7 (Chapter 2)**)[240]. However, as specific HBV mutations for each individual were not consistently listed in each paper, and sequences were typically not available in the public domain for us to review independently, I accepted the conclusion presented in the primary paper regarding the presence or absence of genotypic resistance. For TFV, where there is less clarity regarding the definition of genotypic resistance, for each paper I summarised which RAMs were identified, and whether or not they were reported as resistant (**Appendix 3: Supplementary Table 10**).

Data analysis

I analysed the data in four groups determined *a priori*, under the supervision of Yusuke Shimakawa (Pasteur Institute, France):

(i) "Naive/Entecavir" - previously NA naive individuals treated with ETV,

(ii) "Experienced/Entecavir" - previously NA experienced individuals treated with ETV,

(iii) “Naive/Tenofovir” - previously NA naive individuals treated with TFV,

(iv) “Experienced/Tenofovir” - previously NA experienced individuals treated with TFV.

My primary objective was to estimate the risk of genotypic resistance and associated 95% confidence interval (CI) for all studies in the four groups detailed above after 1yr, 2yr, 3yr, 4yr and ≥5yr of treatment. I grouped time points as follows: 1 year (17 months or less), 2 years (18 - 29 months), 3 years (30 - 41 months), 4 years (42 - 53 months), ≥5 years (54 months and over). Where the same or overlapping datasets were presented in two studies categorised within the same time point, I included the report with the largest sample size. If the sample size was the same, the study with the lowest risk of bias was included. I performed a random-effects meta-analysis of proportions (DerSimonian method with a Freeman Tukey transformation), produced forest plots and pooled risk across studies and calculated the heterogeneity statistic I^2. I assessed heterogeneity arising from clinical and methodological diversity by performing sensitivity analyses where heterogeneity in the primary analysis was present (ie. where pooled estimate was generated and $I^2 \neq 0\%$). My secondary objective, if the number of studies allowed, was to estimate risk of antiviral resistance across subgroups (study design, WHO region, age, proportion of males, HIV coinfection status, HBeAg status, presence of baseline resistance mutations and criteria for sequencing), calculating the antiviral resistance risk and associated 95% CI.

Software

I exported all references to EndNote 20 (Thomson Reuters, New York, NY) and uploaded to Covidence.org. I removed duplicates automatically through Covidence.org and manually if I identified any further duplicates. The team used Covidence software for study selection and data extraction. I performed statistical analyses using the metaprop package in R (version 4.2.2).

5.5.3. Results

Study characteristics

Among 7387 studies identified, 590 full texts were screened for eligibility and 62 (involving 12,358 participants) were included in our analyses (**Figure 5.3**). Of these, 11 studied NA naive individuals subsequently receiving TFV (Naive/Tenofovir), 22 studied NA experienced individuals treated with TFV (Experienced/Tenofovir), 24 studied NA naive individuals treated with ETV (Naive/Entecavir) and 18 studied NA experienced individuals treated with ETV (Experienced/Entecavir) (**Figures 5.3**).

Study characteristics are presented in **Appendix 3: Supplementary Table 11** and **Figure 5.4**. There was an even split of study type; 50% of studies were clinical trials and 50% observational cohort studies (**Figure 5.4 C**). The majority (59/62, 95%) were studies of adults, one studied children only and two included both (**Figure 5.4 D**). Most (54/62, 87%) included a mix of HBeAg positive and negative individuals, one reported data from HBeAg negative individuals and seven from HBeAg positive individuals only (**Figure 5.4 E**). HBV/HIV coinfection was present in two studies, however in general reporting was poor with HIV status not reported in 14/62 (22%) of studies (**Figure 5.4 F**). The global distribution was skewed to the WHO Western Pacific region (49 studies), with only two studies in the WHO African region (**Figure 5.4 G**).

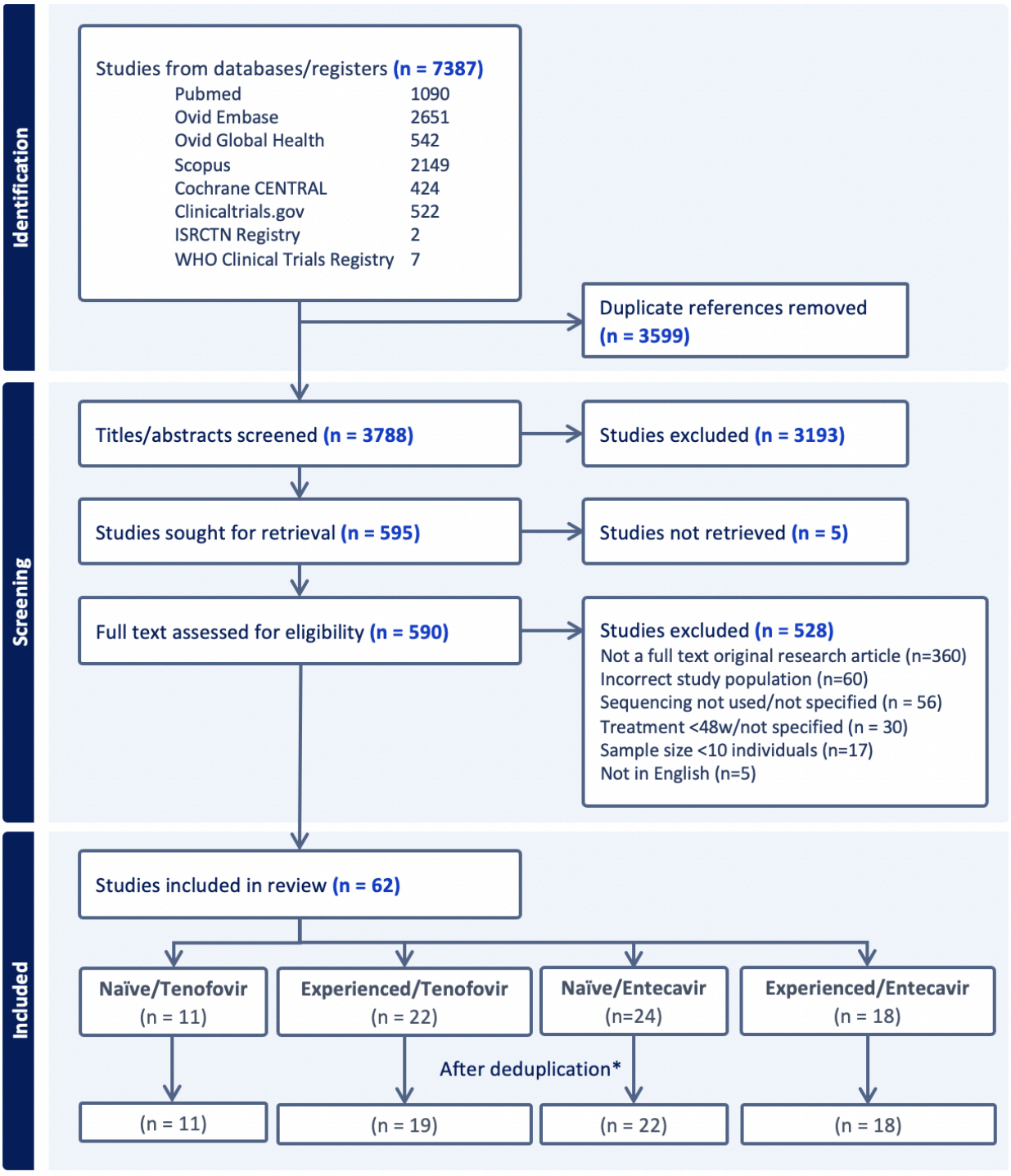


Figure 5.3. PRISMA flowchart for HBVDR systematic review

* For primary analysis, multiple studies reporting the same cohort at the same time point were deduplicated. See materials and methods for details.

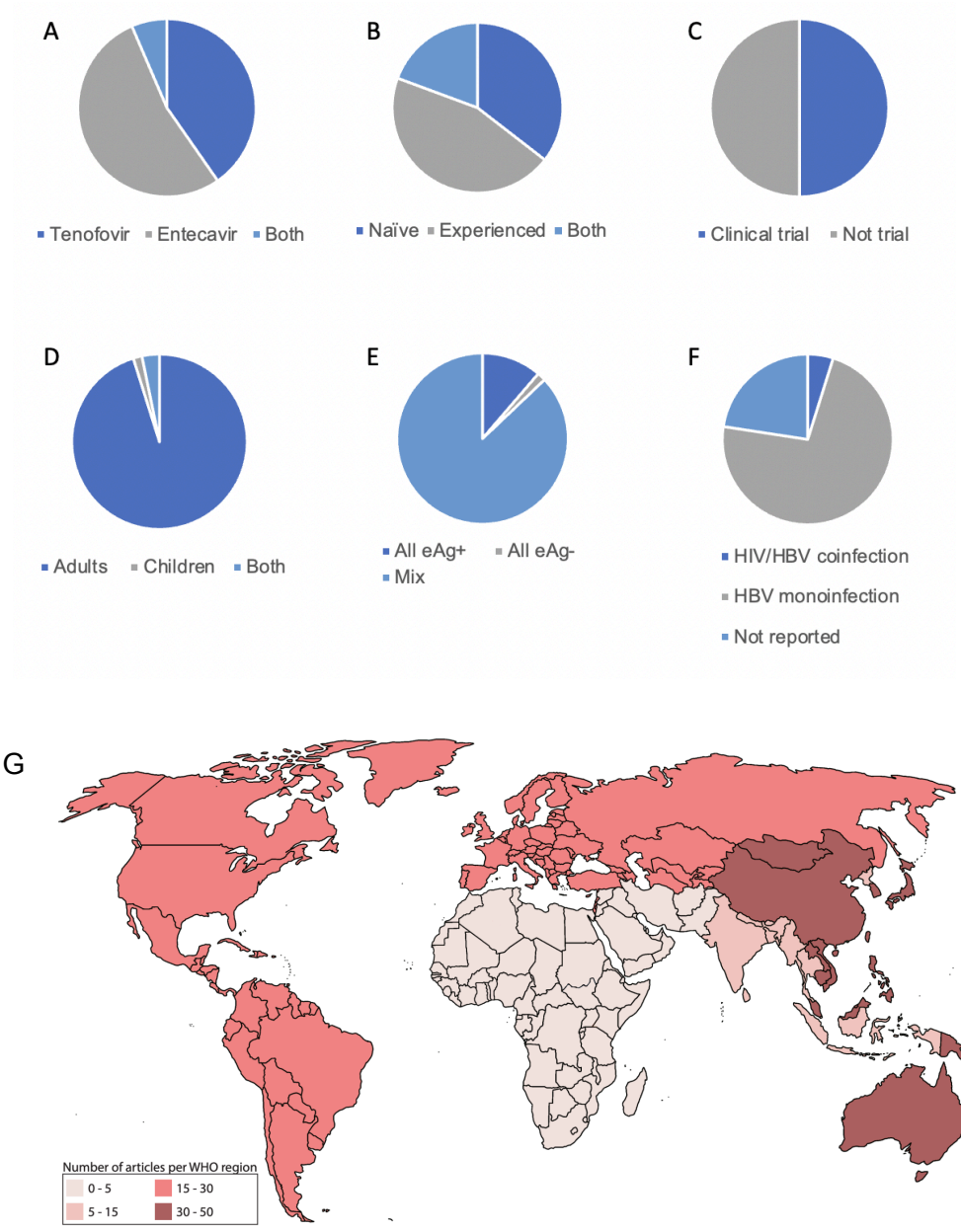


Figure 5.4. Characteristics of 62 studies included in our systematic review and meta analysis to determine the risk of TFV and ETV resistance in individuals receiving treatment for CHB infection. (A) NA being studied, (B) Prior exposure to NA agents, (C) Study type, (D) Population age category, (E) HBeAg status of participants at baseline, (F) Presence of HBV/HIV coinfection, (G) Distribution of studies per WHO regions.

Risk of ETV and TFV resistance over time

The primary objective was to estimate the risk of genotypic resistance in the four main treatment groups (based on prior exposure/current treatment) after 1, 2, 3, 4 and ≥5 years of treatment.

i) Resistance in NA naive individuals treated with ETV

Twenty four studies reported risk of ETV resistance in "naive/entecavir" groups. I excluded two studies as data from the same cohort was published multiple times within the same time point (see Methods section 6.2 for deduplication details). The remaining 22 studies involved 4326 individuals (**Figure 5.5, Appendix 3: Supplementary table 11**). A low risk of resistance that increased with the duration of ETV treatment was seen. After 1 year on ETV, the pooled estimate of resistance was 0.0% (95% CI 0.0 - 0.2%), after 2 years 0.3% (95% CI 0.0 - 0.9%), after 4 years 0.6% (95% CI 0.00 - 0.27%) and at ≥5 years 0.9% (95% CI 0.1 - 2.3%).

ii) Resistance in NA experienced individuals treated with ETV

Eighteen studies involving 1112 individuals reported risk of ETV mutations in "experienced/entecavir" groups (**Figure 5.6, Appendix 3: Supplementary table 11**). ETV resistance was common and increased over time. At 1 year the pooled estimate of resistance was 0.2% (95% CI 0.0 - 1.2%), at 2 years 17.0% (95% CI 5.4 - 32.9%), at 3 years 22.6% (8.2 - 41.0%) and ≥5 years 20.1% (1.6 - 50.1%).

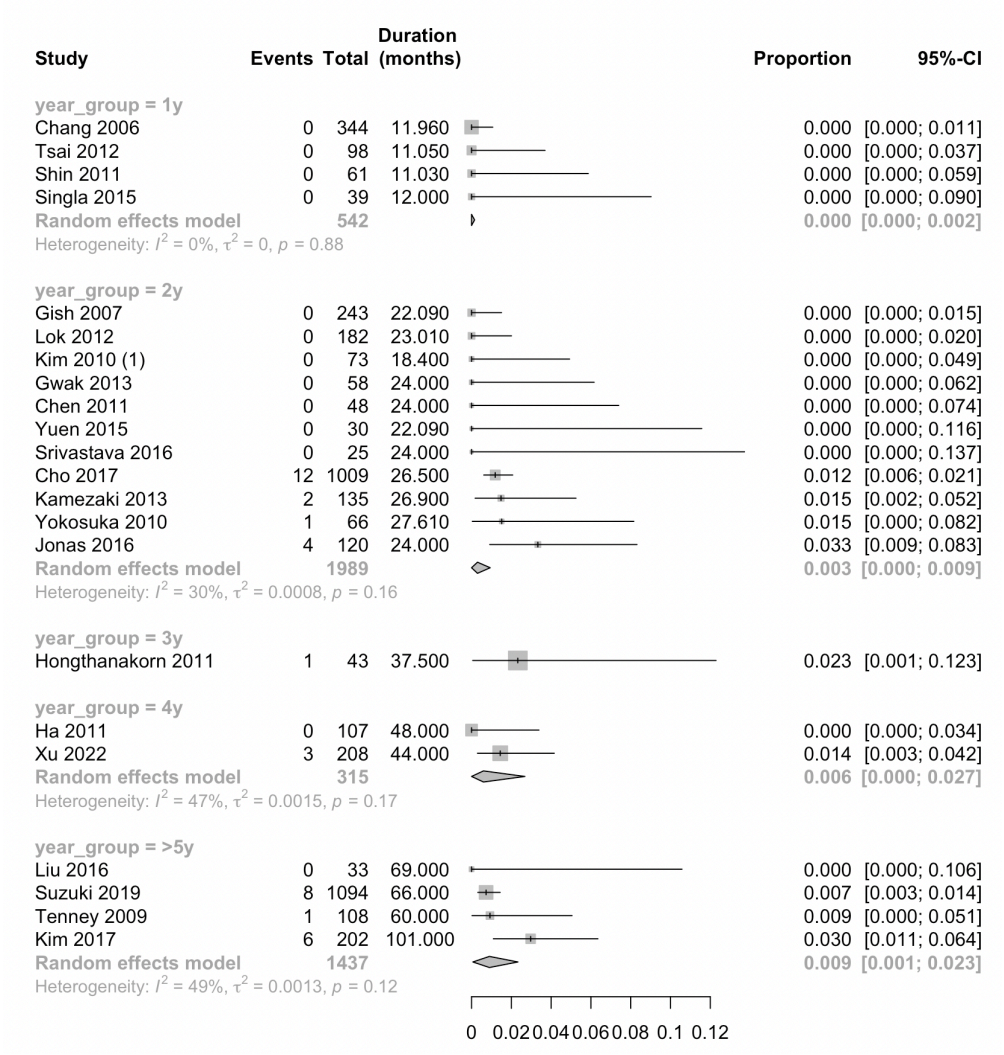


Figure 5.5. Meta-analysis of risk of ETV resistance over time in NA naive individuals.

Forest plot grouped by duration of current NA treatment

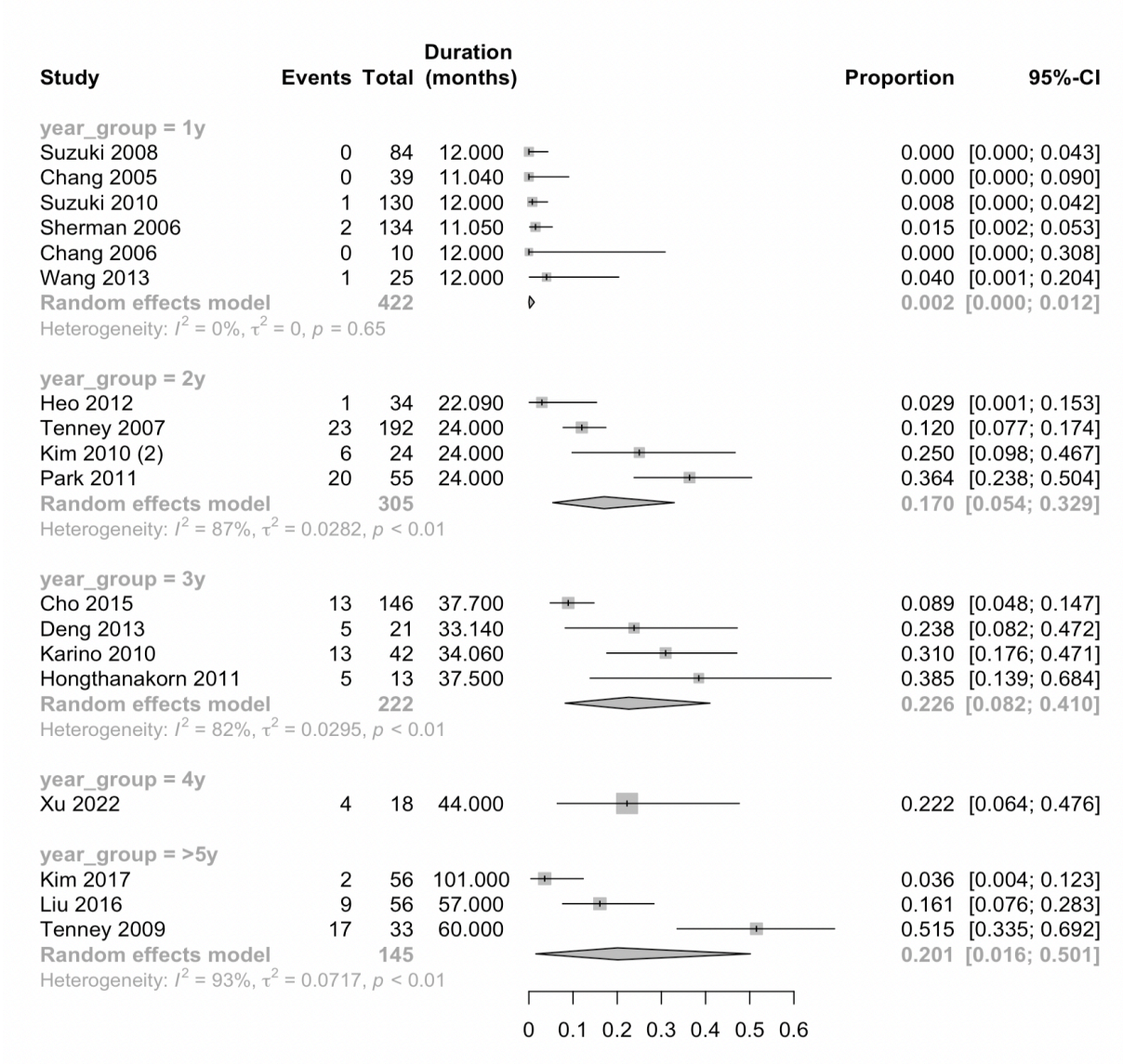


Figure 5.6. Meta-analysis of risk of ETV resistance over time in NA experienced individuals. Forest plot grouped by duration of current NA treatment

iii) Resistance in previously NA naive individuals treated with TFV

Eleven studies including 3778 individuals reported risk of TFV resistance in the "naive/tenofovir" groups (**Figure 5.7, Appendix 3: Supplementary table 11**). Very low risk resistance to TFV was reported, with pooled estimates of 0.0% at all time points (95% CI 0.0% - 0.2% at 1 year, 95% CI 0.0 - 0.0% at 2 years and 95% CI 0.0 - 0.1% at ≥5 years).

One individual in one study[241] was reported to have genotypic TFV resistance. The HBV sequence harboured V173L/V, L180M/L, M204V/M and S213T at baseline, which persisted at week 48 on treatment (putative TFV RAMs, **Appendix 3: Supplementary Table 10**).

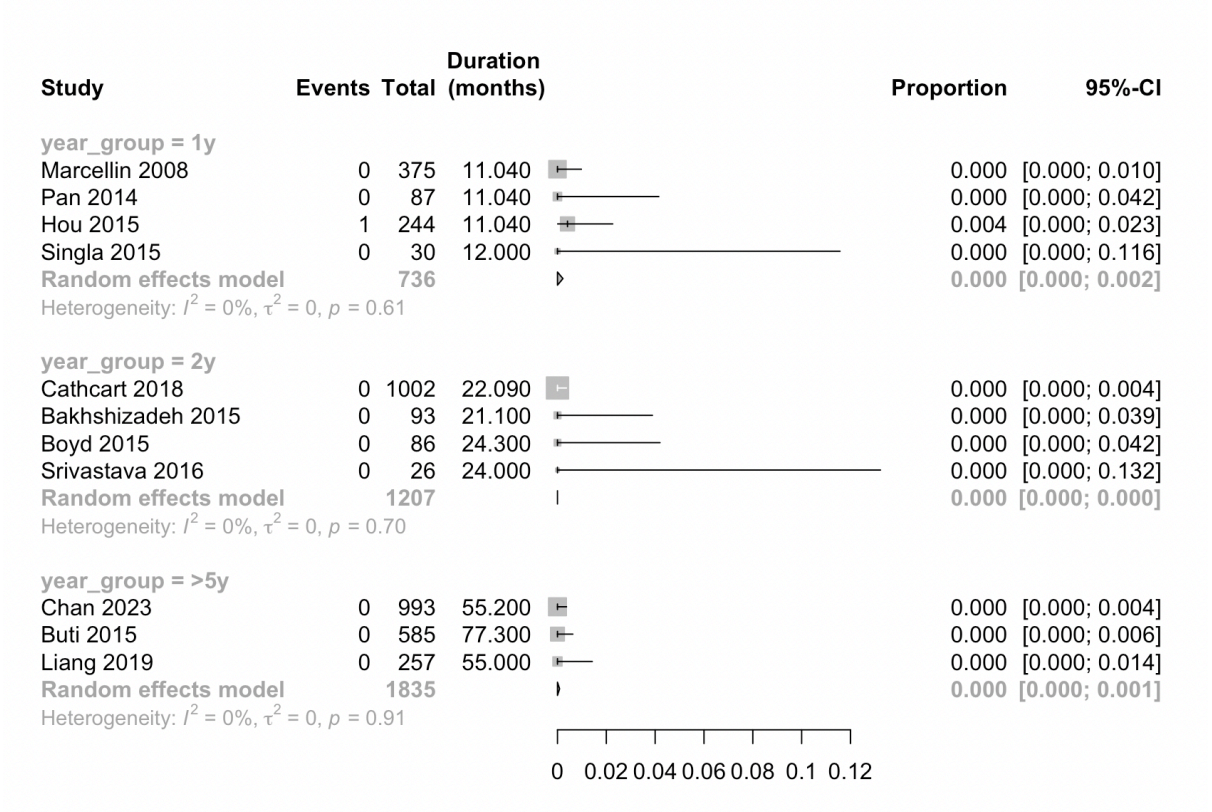


Figure 5.7. Meta-analysis of risk of TFV resistance over time in NA naive individuals. Forest plot grouped by duration of current NA treatment

iv) Resistance in NA experienced individuals treated with TFV

Twenty two studies reported risk of TFV mutations in the "experienced/tenofovir" group. Three studies were excluded from meta-analysis as data from the same cohort were published multiple times within the same time point (see Methods for details of deduplication). The remaining 19 studies involved 2059 individuals (**Figure 5.8, Appendix 3: Supplementary table 11**). All pooled estimates were 0.0%, with 95% confidence intervals of 0.0 - 0.7% at 1 year, 0.0 - 0.6% at 2 years, 0.0 - 2.2% at 3 years and 0.0 - 0.0% at ≥5 years.

Although the pooled estimate for tenofovir resistance were 0.0%, TFV resistance was reported in five individuals across three studies (**Figure 5.8, Appendix 3: Supplementary table 11**). In one small study performed in Australia 2006-2008[242], 3/38 individuals had evidence of resistance, two with an N236T mutation at baseline persisting on TDF, and one developing A181T and N236T on treatment (established TFV RAMs, **Table 2.7 (Chapter 2)**). A cohort study in the USA performed between 1999-2007[243], classified 1/10 individuals as having genotypic resistance to TFV. The individual had received prior treatment with adefovir and lamivudine and had adefovir resistance at baseline (A181V on direct sequencing and N236T on clonal analysis) which persisted during TDF treatment (established TFV RAMs, **Table 2.7 (Chapter 2)**). In a retrospective study performed in 2012 in Italy[244], 1/11 individuals were classified as having TFV resistance. The HBV VL did not respond to prior lamivudine and adefovir treatment. Although relevant RAMs were not detected at baseline, on follow-up following TFV treatment an A194T mutation was identified (putative TFV RAM **Table 2.7 (Chapter 2)**). Individuals with incomplete adherence were not excluded from the latter two studies.

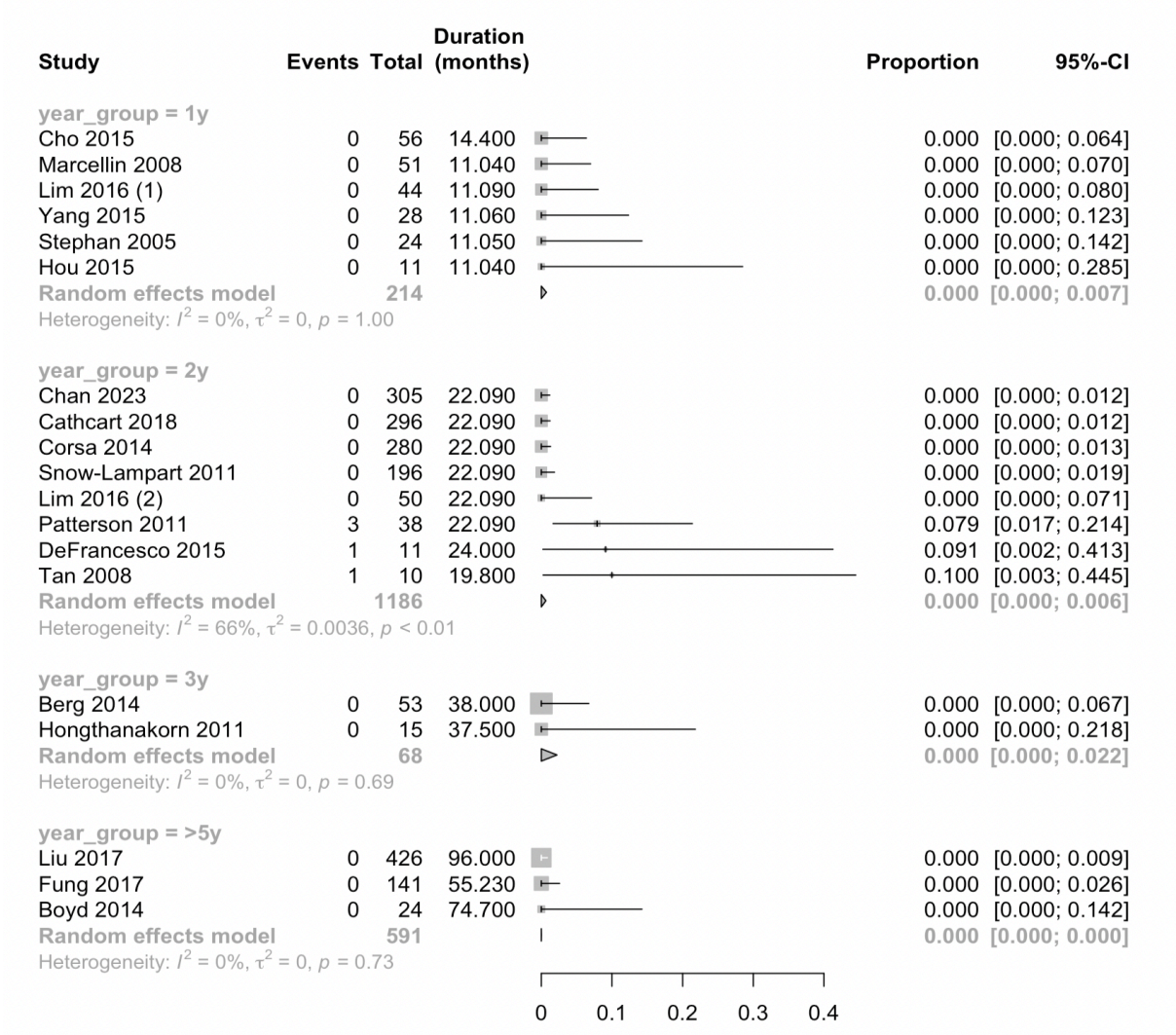


Figure 5.8. Meta-analysis of risk of TFV resistance over time in NA experienced individuals. Forest plot grouped by duration of current NA treatment

Subgroup, sensitivity and outlier analysis

Due to the small number of studies within each exposure/treatment/year group category, and variable reporting of age/sex, subgroup analysis was not performed. I explored study heterogeneity arising from clinical and methodological diversity by performing sensitivity analyses where heterogeneity in the primary analysis was present (ie. where pooled estimate was generated and $I^2 \neq 0\%$) and also performed an in depth assessment of outliers. I did not identify any variables consistently affecting the heterogeneity of estimates.

In NA naive individuals on TFV, heterogeneity (I^2) was 0% at all time points, therefore I did not perform sensitivity analyses. Only one individual in one study[241] was reported to have genotypic TFV resistance. This individual was reported as nucleos(t)ide naive, but the HBV sequence harboured V173L/V, L180M/L, M204V/M and S213T at baseline, which persisted at week 48 on treatment (putative TFV RAMs **Table 2.7 (Chapter 2)**).

In NA experienced individuals treated with TFV, heterogeneity was only seen at the 2 year time point ($I^2 = 66\%$). Restricting studies to those where all individuals were sequenced led to the largest reduction in heterogeneity to 4%, restricting by study size reduced I^2 to 56%, by study design to 56% and by excluding those where all had baseline RAMs to other NAs to 39%, implying multiple factors impacted on heterogeneity (Appendix 3: Supplementary Table 4). One study[242] was classified as an outlier, with its 95% CI not overlapping with the CI of the pooled estimate. It was a small study performed in Australia 2006-2008, 3/38 individuals developed evidence of genotypic resistance, two individuals had N236T mutation at baseline which persisted on TFV treatment, one individual developed A181T and N236T on treatment (accepted TFV RAMs, **Table 2.7** (**Chapter 2**))[242].

Reported TFV resistance was a rare event, therefore the remaining studies reporting resistance (2/20) were investigated as potential outliers. On detailed analysis, both studies were small (n=10,11) with prior ADV +/- 3TC exposure. A cohort study in the USA performed between 1999-2007[243], classified 1/10 individuals as having genotypic resistance to TFV. The individual had received prior treatment with ADV and 3TC and had ADV resistance at baseline (A181V on direct sequencing and N236T on clonal analysis) which persisted during TFV treatment (accepted TFV RAMs, Table 2.7 (**Chapter 2**)). In a retrospective study performed in 2012 in Italy[244], 1/11 individuals were classified as having genotypic resistance to TFV. This individual did not respond to prior 3TC and ADV treatment however did not have 3TC or ADV RAMs at baseline. They developed an A194T mutation on TFV (putative TFV RAM Table 2.7 (**Chapter 2**)). Individuals with incomplete adherence were not excluded from any of these two studies.

In order to establish whether heterogeneity in the estimate for risk of TFV resistance could have been influenced by variations in thresholds for calling TFV RAMs, I compared the resistance criteria across papers (**Table 2.7** (**Chapter 2**)). Among the 26 papers included in the primary analysis for TFV (irrespective of whether previously naive or treatment experienced), only five classified mutations A181T/V and N236T known to be linked to reduced TFV sensitivity as RAMs, seven papers identified these mutations but did not classify them as resistant. Three of the 26 papers reported one or more of the putative mutations listed in Table 2.7 (**Chapter 2**) as RAMs, an additional 8 papers identified one or more of these combinations of mutations but did not report the individual as resistant.

In previously NA-naive individuals prospectively followed up after starting ETV, heterogeneity was seen at years 2, 4 and ≥5 time points. None of the sensitivity analyses performed reduced the heterogeneity (**Appendix 3: Supplementary Table 5**).

On analysis of heterogeneity of studies at year 2, the highest resistance risk estimate at 3.3% was a multi-centre clinical trial of HBeAg positive children, all individuals developed M204M/V, L180M, S202G mutations (the only other study recruiting children [245] was also in the same subgroup with a resistance risk of 0.0%). At years 4 and ≥5, both papers reporting no resistance used more stringent criteria for sequencing, only investigating individuals meeting the criteria for VBT, rather than all those with detectable viraemia. This may lead to an under-estimation of primary non-response to ETV in naive individuals. A retrospective cohort study in Korea[93] reported the highest risk of 3% at ≥5 years. The duration of follow up was longer here than the other trials in the same time bracket (101 vs 60-70 months) which may explain the higher proportion of resistance observed (however the same study also reported on previously NA experienced individuals, and a relatively low risk of resistance was identified).

In previously NA experienced individuals on ETV, heterogeneity was seen at 2, 3 and 5+ year timepoints. Sensitivity analysis only showed a reduction in heterogeneity when restricting based on sequencing criteria at year 2 (reducing heterogeneity from 97% to 62%) and presence of baseline RAMs at year 5, reducing heterogeneity from 93% to 80%. **Supplementary Figure 2** (**Appendix**) shows the impact of baseline RAMs across all papers, those studies excluding RAMs reported a lower risk of ETV resistance (<10%) than those where some or all individuals had RAMs at baseline (up to 51.5% resistance).

On analysis of the highest/lowest risk at each time point, at 2 years, the study with the lowest resistance risk, a clinical trial in Korea (2.9% resistance) excluded those with baseline RAMs but sequenced all individuals with detectable VL[246]. The study with the highest estimate of 36.4% studied individuals all of whom had genotypic resistance to lamivudine at baseline and

were also refractory to adefovir[88]. The difference in baseline populations may contribute to heterogeneity here due to the influence of cross-resistance.

At 3 years, the lowest estimate of 8.9% was a retrospective study of 146 individuals without baseline mutations[247]. The highest estimate (38.5%) came from a retrospective cohort of 13 individuals with detectable VL on treatment[89]. Details of prior NA exposure was not reported for this subgroup and baseline RAMs were not reported; only those with VBT were sequenced, however a line probe assay was used in addition to sequencing to determine genotypic resistance, and it is not clear whether the results presented are from sequencing alone, or the line probe assay.

The lowest estimate of resistance risk at ≥5 years was a retrospective cohort study in Korea[93] (3.6% ETV resistance), which excluded individuals with any RAMs at baseline and only performed sequencing on individuals with VBT. In contrast, in the study with the highest estimate (51.5% ETV resistance) all individuals had RAMs to other NAs at baseline and on follow up sequenced all individuals with a detectable VL[91].

Four studies[93,248–250] used a lower dose of ETV (0.5mg OD) than the 1.0mg dose currently licensed for cases of resistance (where studies used multiple doses, data from the 1mg OD group was extracted). However it was not possible to assess the impact of dosing on heterogeneity as the majority of these studies were in a year group where no heterogeneity was observed. All studies except one[248] included individuals with prior 3TC experience, so it was not possible to determine whether prior 3TC vs. other NA was a risk factor for ETV resistance.

Quality and risk of bias

In the primary analysis, distribution of studies of low/moderate risk was consistent across the naive/tenofovir (5 low:6 moderate) experienced/tenofovir (8 low:10 moderate:1 high), naive/entecavir (10 low:12 moderate) subgroups. In the experienced/entecavir group, a higher proportion of moderate risk studies was seen (6 low:11 moderate:1 high). The quality/risk of bias scores for individual studies are presented in **Appendix 3: Supplementary Table 11** and **Supplementary Figure 1**).

Criteria for inclusion were well reported in most studies (87%) included in the primary analysis, however there was clinical heterogeneity between the populations included, particularly within the NA experienced cohorts. Inclusion criteria ranged from individuals with prior NA treatment without clinical/genotypic resistance, to populations where all had clinical but not genotypic resistance and others where all had clinical/genotypic resistance (with RAMs to other NA) before starting TFV/ETV. All studies used objective standard criteria for the diagnosis of CHB.

Study subjects were described in detail in 82% of the studies included in the primary analysis. However in some studies, demographic characteristics at baseline were not provided at the individual level, meaning subgroup analyses based on age and sex could not be performed. There was variable reporting of suspected risk factors for resistance e.g. HIV status, HBeAg status, drug adherence and genotype of infection (which is of importance, as some genotypes have certain RAMs as wild type consensus[86]). The duration of TFV/ETV treatment spanned a wider range in cohort and cross-sectional studies compared to clinical trials where all individuals were followed up for the same duration, potentially providing a less accurate estimate of the relationship between duration of therapy and risk of resistance. There was sometimes a lack of clarity about prior agents used, duration of prior therapy and breaks in therapy. We were unable to consistently extract follow up data (**Figure 5.2**); for

consistency, denominators were recorded at baseline, however in some cohorts some individuals switched treatment for undefined reasons, there was lack of clarity regarding which individuals had HBV sequencing undertaken on cessation/switching therapy, and in some cases individuals were lost to follow up.

Outcomes were measured in a valid way in 92% of studies. However, a range of sequencing methods were used (from those reporting consensus sequences to quasispecies), with varying sensitivity for detecting resistance. Criteria for sequencing differed between studies, some sequencing all viraemic individuals but others only sequencing those with VBT, risking underestimating the risk of resistance. Definitions of drug resistance were inconsistent between studies of TFV (**Supplementary table 10** (**Appendix 3**), **Table 2.7** (**Chapter 2**)). In some papers, the criteria used to determine resistance were not specified, and when they were, studies used different thresholds/criteria to call resistance. Since not all polymorphisms were listed for each individual and the majority of sequences were not published, I had to use the authors' judgement on what constituted genotypic resistance, rather than being able to take an objective view using the same criteria across all studies (ie. a predefined list of relevant RAM combinations). Furthermore, sequencing was not always successful, which may lead to an under-estimation of resistance risk. The relationship between clinical and methodological sources of heterogeneity and the observed risk of genotypic resistance over time is presented in **Appendix 3: Supplementary Figure 2**.

Due to the decision made a priori to present results of NA naive and NA experienced individuals separately, some papers were excluded as resistance risk was not reported for naive and experienced subgroups separately, or the denominator was not clear and resistance risk could not be calculated[251–257].

Publication bias was assessed with funnel plots of study size against log odds. For TFV, most studies reported zero events and therefore small-study effects could not be assessed. For ETV, the adapted funnel plot was asymmetrical for studies of individuals who were previously NA naive and symmetrical for studies of individuals who were previously NA experienced, indicating that there might be a publication bias for studies of NA naive individuals, towards smaller studies reporting a higher risk (**Appendix 3: Supplementary Figure 3**).

5.5.4 Discussion

These data confirm a higher risk of genotypic resistance with ETV treatment, particularly in patients with prior failure (and/or confirmed resistance) to other NAs. I have been able to calculate a pooled estimate for the first time, quantifying this as 20.1% over 5 years (CI 1.6 - 50.1) in the group with prior NA exposure. TFV has an excellent resistance profile in both treatment naive and experienced individuals, even with long term therapy, confirming its high genetic barrier to the development of resistance. However, I note some studies that did identify TFV resistance[241–244] and highlight important gaps in existing data and challenges with interpreting TFV resistance data, due to variation of resistance calling algorithms and definitions in use. This evidence reaffirms the current protocols for use of TFV as the first line NA in patients with prior failure of or resistance to other agents, but with a need for ongoing scrutiny[57,72,74,258].

This meta-analysis was limited by the small number of trials within each group. Clinical and methodological heterogeneity was observed (see Quality section for discussion), with variation in inclusion criteria (including presence/absence of baseline RAMs), definitions of partial response, thresholds for sequencing, and varied definitions of TFV drug resistance potentially impacting on resistance risk estimates.

In order to optimise the quality and reduce bias, I excluded smaller studies and those only focusing on drug-resistant populations, and did not include studies that did not present viral sequence data. This focused approach means I have not accounted for the full spectrum of published data on drug resistance; assimilation of evidence from studies specifically focusing on drug resistant populations could be a future aspiration to collect data on risk factors, clinical characteristics and sequence motifs associated with resistance. Future studies should aim to identify novel as well as known RAMs. Challenges around representative data are discussed below.

5.6. WHO stakeholder analysis: knowledge gaps and plan for the field

Here I report outputs from a multidisciplinary consultation convened by the WHO in 2023 to discuss HBVDR. I have already discussed the key HBVDR concepts and definitions from this manuscript in **Chapter 1**, **sections 1.3.10 - 1.3.13** and summarise key knowledge gaps and a research roadmap here.

5.6.1 Knowledge gaps

There are a number of sources of bias and knowledge gaps that impede insights into HBVDR. Most HBVDR data come from clinical trials involving well-monitored patients in resource-rich settings, often excluding those with complex health issues (like the individual in my case study above) or unsuppressed VLs. Key populations, such as people with mental health challenges, LGBTQIA+ individuals, migrants, those experiencing housing instability, and people who inject drugs, are underrepresented. Additionally, resource-limited regions with high HBV prevalence are poorly represented in clinical trials and cohorts, leaving gaps in understanding HBVDR across diverse populations, genotypes, and co-infections (e.g. HIV or HDV).

HBV sequencing is not routinely performed before treatment initiation in routine clinical practice, and even when generated sequences are often not shared publicly, making it hard to assess ADR and TDR prevalence. In the future, reflex sequencing could be considered, where residual plasma from individuals with detectable HBV VL are earmarked for sequencing. Sequencing typically succeeds only in cases with high VLs, leaving a knowledge gap about RAMs in low-level viraemia (gap narrowed by my HEP-TILE method, **Chapter 4**). Sequencing methods for resistance determination may include whole genome (e.g.

HEP-TILE), or partial genome sequencing, using a shorter amplicon that spans the Spacer/RT domains and overlapping S gene (for the detection of immune and vaccine escape). A 450bp region in the spacer domain has been shown to be sufficient for genotyping (unpublished data from Matthews group). The use of sub-genome fragments has the advantage of improving ease and sensitivity of sequencing workflows, however has the disadvantage of missing a proportion of recombinant infections and other clinically relevant polymorphisms in the pre-core/core and X genes. Metadata for viral sequence data often lacks crucial treatment history, details on adherence or clinical outcomes (e.g. liver cirrhosis and HCC), limiting analysis of specific resistance risk factors and their clinical impact.

5.6.2 Closing the research and data gaps in the HBVDR field

There are a number of domains that need to be tackled in parallel to close the current HBVDR data gap and mitigate against development of HBVDR, which I present in **Table 5.4**. Expanding HBVDR testing using infrastructures built for HIV and SARS-CoV-2 can enhance data consistency and access to sequencing resources. Development of integrated programmes will enhance access to laboratory testing, unify reporting of data, reduce inequities, improve efficiency, and potentially offer cost-savings, as resources and expertise are pooled. This work needs to happen in parallel with efforts to optimise the care cascade. Optimising treatment availability requires robust infrastructure, including preventing pharmacy stock-outs, considering multi-month dispensing, minimising out-of-pocket costs, provision of counselling and education, and monitoring of adherence (e.g. self-reporting, VL monitoring, refill records).

Table 5.4. Suggestions for closing research gaps in the HBVDR field. This table summarises outputs from the WHO HBVDR meeting in 2023, subsequently refined based on expert opinion from an interdisciplinary stakeholder group.

Domain	Suggested approaches to close data gaps
Defining and monitoring therapeutic response	**Definitions and data** > Unify consensus definitions of primary response, partial response and virologic breakthrough for ETV and TFV. > Establish agent-specific predictors and risk factors for suboptimal treatment response in diverse populations. > Evaluate agent-specific prevalence of suboptimal treatment response and drug resistance. > Establish longitudinal databases with accurate treatment histories. **Clinical impact** > Evaluate the impact of HBVDR to ETV and TFV on outcomes such as histological improvement of liver disease, development of further drug resistance and adverse events. > Evaluate the impact of HBVDR minority variants on clinical and treatment outcomes and determine clinically relevant thresholds for drug resistant minority variants. > Evaluate the transmissibility and viral replicative fitness costs of HBV RAMs. > Determine the impact of treatment adherence ± drug levels on virologic suppression and presence of RAMs. **Monitoring** > Evaluate the relative impact and cost-effectiveness of a range of HBV DNA quantification schedules for the detection of treatment failure and drug resistance. > Evaluate lower cost and point of care tests for HBV DNA VL quantification as potential markers for treatment failure and drug resistance. > Evaluate different testing substrates, e.g. dried blood spots > Evaluate the utility and predictive value of monitoring ALT levels and other markers to identify the development of genotypic or phenotypic resistance. > Establish therapeutic windows for drug levels of NAs for the treatment of CHB, identifying any differences in therapeutic drug concentrations for HIV and HBV.
Clinical interventions	> Establish the optimal clinical care for persons with primary non response, persistent viraemia and VBT on antiviral monotherapy, with and without confirmed RAMs, based on real-world data, which may require pooled longitudinal multi-centre clinical and virological data > Establish a consensus around agent-specific time points at which treatment ‘failure’ should be diagnosed and acted upon > Determine cost-effectiveness of different antiviral strategies (e.g. switching agent vs. adding agents) > Update clinical guidelines based on evidence of best practice for identification and management of drug resistance for all HBV active agents
HBVDR testing	> Consider leveraging existing WHO HIV drug resistance laboratory network to establish an HBV/integrated viral drug resistance laboratory network to support HBV VL monitoring, resistance testing and collate existing HBVDR data > Develop an HBVDR laboratory operational framework, including guidance for genotyping

	> Develop a target product profile for HBVDR testing to guide development and evaluation of laboratory protocols > Increase investment in deep sequencing and full genome sequencing for HBV to develop low cost, higher sensitivity HBV sequencing protocols > Establish a validated open-access genotyping interpretation algorithm based on currently available tools.
Quality of care indicators and care cascade	> Evaluate and address barriers to diagnosis, linkage to care, retention to care and treatment adherence > Improve advocacy and peer-support to increase adherence > Evaluate acceptability and efficacy of integrated care pathways with holistic client-centred care > Expand the WHO Global Viral Hepatitis Reporting Framework to include quality of care indicators to track the cascade of care > Establish evidence based early warning indicators for HBVDR > Perform implementation research - management of resistance (or treatment failure) in centralised vs decentralised services (clinical, labs, data collection) > Learn from experiences of existing HIVDR and HBVDR programmes
Awareness, education and representation	> Involve the community and civil society to improve knowledge and education about drug therapy, and to design patient-centric research > Engage with communities and people with lived experience of HBV infection to establish the extent to which knowledge and discussion of drug resistance is helpful and/or desirable; > Increase awareness of drug resistance among diverse stakeholders (clinicians, laboratory staff, research community, public health, policy-makers), and community groups.
Representative clinical data	> Identify key groups for HBVDR surveillance, including at population-level and specific high-risk groups > Conduct longitudinal cohort studies especially in populations in the WHO African region, but also in under-researched populations, to evaluate the prevalence of and risk factors for HBVDR > Include surveillance for HBVDR in protocols for treatment studies > Use high quality data to model the impact of drug resistance over time on individual and population suppression of viraemia and on transmission rates.
Genomic data publication and data sharing	> Collect, analyse and publish viral sequence alongside clinical metadata that includes: **> *essential*** data (unique de-identified person ID, HBV DNA VL antiviral agent(s), duration and dates, country); **> *desirable*** data (specimen type, method of sequencing, coinfection, cirrhosis/HCC and other comorbidity, drug levels, *in vitro* susceptibility data, virological response to NA regimens). > Establish centralised open-access repositories and resources for collation of international data > Establish approaches to ethical international data sharing, with rigorous data governance and archiving > Clear reporting of which individual/combinations of RAMs were identified and considered significant.

HBVDR – Hepatitis B virus drug resistance; NA – nucleos(t)ide analogue; RAM – resistance associated mutation; HIV – human immunodeficiency virus; VL – viral load; ETV – entecavir; TFV – tenofovir, VBT = viral breakthrough, POCT – point of care testing

5.7. Conclusion and future work

My assessment of whether treatment failure and HBVDR is a threat to HBV elimination globally is currently compromised by inadequate data. In this chapter I have shown that treatment failure can be underpinned by diverse inter-related factors, one of which is HBVDR. In a meta analysis of published data, I quantified the prevalence of resistance to TVF and ETV in treatment-naïve and exposed patients, but recognise that this is a biased picture of the true landscape due to the poor representation of the global landscape, and lack of inclusion of individuals in many groups who have risk factors for treatment failure and drug resistance.

Undertaking a systematic analysis has provided insights into challenges in the current literature, and highlights important avenues for future scrutiny, with immediate translational implications as more people become eligible for, and can access, NA therapy. Future studies should strive to address global representation, to generate high quality, long duration, real world data that includes populations with risk factors for resistance. Standardised reporting of drug resistance, publication of clinical metadata and sharing of sequence data are essential attributes of reporting. These data are essential to guide strategies for implementation of effective therapy that will support progress towards HBV elimination worldwide.

Chapter 6: Treatment eligibility and drug resistance in the WHO African region

6.1. Availability of content

I presented data on transient elastography scores from the ULiDS cohort in a poster at the Conference of Liver Disease in Africa (COLDA) meeting 2023[259]. I presented data regarding treatment eligibility in an oral presentation at the UK HBV meeting in 2024, manuscript in preparation. I presented WGS findings from DRC in two posters at the European Association for the Study of the Liver (EASL) conference 2024, I was first author of one poster[211], and co-author of the other[260,261].

6.2. Background

HBV is highly prevalent in the WHO African region with over 64 million people living with CHB[29]. Only 4.2% individuals with CHB have been diagnosed and 0.2% are receiving treatment, falling short of the WHO 90% testing and 80% treatment coverage targets. To address the low proportion of individuals on treatment in resource-constrained settings, the WHO expanded and simplified HBV treatment guidelines in 2024 (**Figure 6.1**). To implement these guidelines, there is a critical need to understand (1) current treatment eligibility and (2) HBVDR prevalence, to understand the real world implications and inform national treatment programmes. This chapter aims to bridge these two data gaps across four settings. The Matthews research group has team members based in South Africa and Kenya and collaborators based in Uganda and the Democratic Republic of the Congo (DRC) (**Figure 6.2,** see **Chapter 2: Methods** for full description of settings) which led to the work described below.

6.3. Aims

Aim 6.1. To evaluate HBV antiviral treatment eligibility in rural Uganda under the new 2024 WHO guidelines

Aim 6.2. To assess the utility of different non-invasive tests of liver fibrosis/cirrhosis for determining HBV treatment eligibility

Aim 6.3. To use HBV WGS to characterise genotypes, resistance associated mutations, immune escape and transmission across four settings in the WHO African region

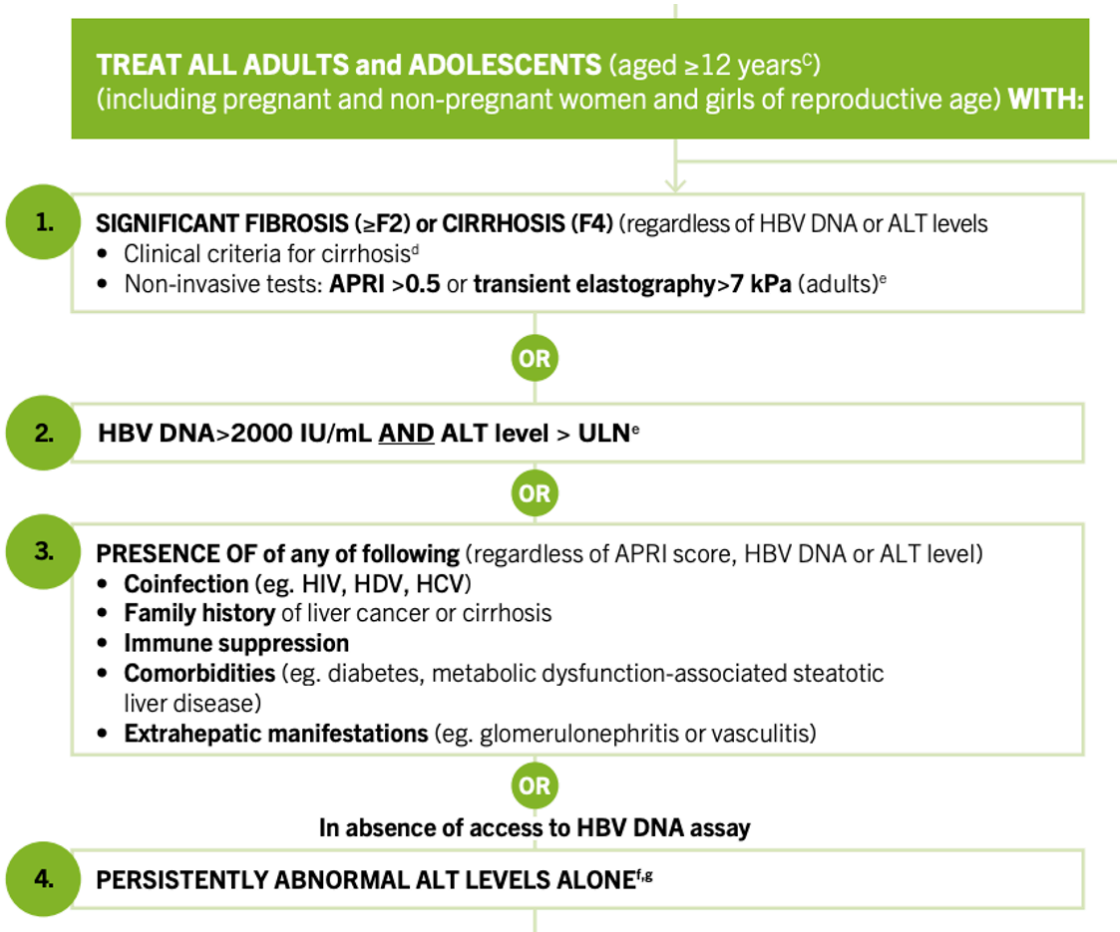


Figure 6.1. HBV treatment eligibility criteria from the WHO guidelines for the prevention, diagnosis, care and treatment for people with chronic hepatitis B infection, March 2024. Reproduced from WHO guidance under a CC BY-NC-SA 3.0 IGO license[262]

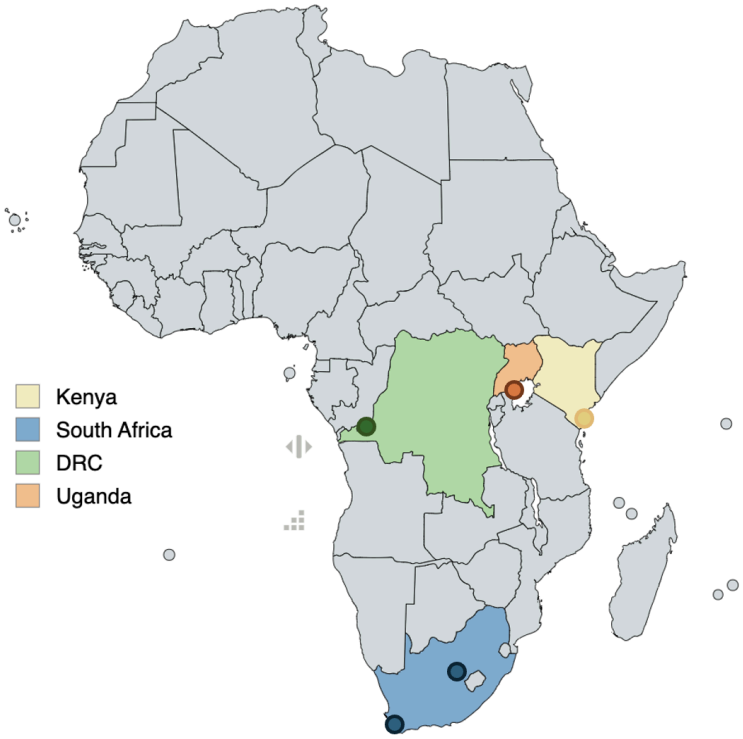


Figure 6.2. Map of Africa showing locations of study settings in Kenya, South Africa, Democratic Republic of the Congo and Uganda. Locations of cohorts within each country shown by coloured dots; yellow - Kilifi (Kenya); orange - Kalungu district (Uganda); green - Kinshasa (DRC), blue - Cape Town and Bloemfontein (South Africa). Produced using MapChart[263]

6.4. HBV treatment eligibility in the Uganda General Population Cohort

6.4.1 Introduction

The 2024 WHO HBV treatment guidelines[262] present four options for determining treatment eligibility (**Table 6.1, Figure 6.2**) based on the presence of liver fibrosis/cirrhosis (APRI or TE), hepatic inflammation (ALT), HBV replication (HBV VL) and past medical and family history. The guidance provides simplified criteria when access to laboratory infrastructure and specialised tests is limited, including the option for use of APRI (with a revised lower threshold) as a marker for fibrosis and treating based on abnormal ALT alone if VL testing is inaccessible.

In 2011, a seroprevalence study nested within the Uganda General Population Cohort (GPC)[123,124] identified 220 individuals with positive HBsAg. These individuals were followed for 12 years, attending for repeat clinical assessment, blood tests and transient liver elastography (TE) in 2023. Here I assess treatment eligibility among a cohort of 63 individuals with confirmed CHB, calculating the proportion of individuals eligible for treatment in a well resourced setting with all tests available, compared to a hypothetical scenario of a less resourced setting without TE.

I compare the sensitivity of three non-invasive tests of liver disease for predicting which individuals have fibrosis/cirrhosis (TE >7kPa) and are therefore eligible for treatment:

i) the WHO APRI score threshold of >0.5 (a “treat-with” approach),

ii) the “treat-all-except” APRI threshold of <0.36, an approach proposed by the HEPSANET consortium[264], where all individuals are treated except those meeting criteria that suggest low risk of disease progression,

iii) the Gamma-Glutamyl Transpeptidase-to-Platelet Ratio (GPR), another non-invasive test (NIT) initially designed for HBV mono-infected individuals in a West African setting[265].

I compare this to a “treat-all” approach, allowing same-day initiation of antiviral therapy without costly clinical staging[266].

Table 6.1. WHO HBV treatment criteria 2024 mapped against data available in Uganda Liver Disease Study[262]. A summary of the treatment guidance and the parameters available in this cohort used to determine treatment eligibility.

WHO treatment criteria		**Data available in ULiDs study**
1	Treat all with significant fibrosis on non-invasive testing (APRI score >0.5 or transient elastography >7KPa), or clinical evidence of cirrhosis	- Clinical evidence of cirrhosis 2023 - APRI 2023 - TE 2023
2	HBV VL > 2000 IU/ml and an ALT level above the upper limit of normal (ULN) (30 U/L for men and boys and 19 U/L for women and girls)	- HBV VL and ALT in 2023 for all - HBV VL and ALT in 2011 if available
3	Treat all with coinfections (such as HIV, hepatitis D or hepatitis C); family history of liver cancer or cirrhosis; comorbidities (such as diabetes or metabolic dysfunction–associated steatotic liver disease) and extrahepatic manifestations (e.g.glomerulonephritis or vasculitis)	- HIV 2011 and 2023 - HCV 2011 - Family history of death by liver disease - Past medical history of diabetes or HbA1c >6.5
4	Two ALT values above the ULN	- ALT in 2011 and 2023

ALT - alanine aminotransferase; APRI - aspartate aminotransferase to platelet ratio index; TE - transient elastography; HBV - Hepatitis B virus; HCV - Hepatitis C virus; HIV - human immunodeficiency virus; VL - viral load; ULN - upper limit of normal.

6.4.2 Methods

Uganda General Population Cohort and Uganda Liver Disease Study (ULiDS) sub-study

The Uganda GPC was established in 1989 to study HIV in rural south-western Uganda[124] with funding from the UK Medical Research Council (MRC) in collaboration with the Uganda Virus Research Institute (UVRI). Since 2010, the scope of research questions has broadened to incorporate the epidemiology and genetics of both communicable and non-communicable diseases. The cohort now consists of ~22,000 individuals living in 25 villages who are followed up on an annual basis.

In 2011, a serosurvey for HIV, HBV and HCV infection was performed alongside liver function tests, lipid profile and assessment of cardiometabolic risk factors for 8099 individuals over the age of 13 years old (results previously reported[123]) (**Figure 6.3**). This identified 220 people with positive HBsAg. These 220 individuals were re-surveyed in 2023, as part of ULiDS, with repeat testing as described in **Figure 6.3**. All individuals in the ULiDs sub-study were born prior to the introduction of HepB3 under the Uganda National Expanded Programme on Immunization (UNEPI) in 2002. Antiretroviral therapy with an HBV active agent was available to those with HIV/HBV coinfection after diagnosis.

Non-invasive testing for liver disease was performed in 2023 with an APRI score (APRI = (AST level/40)/platelet count (10^9/L) x100) and fasting TE (Fibroscan™, (Echosens Paris) with QC criteria of >= 10 measurements and IQR/Med < 30%) and GPR score (GPR = GGT/platelet count (10^9/L)). Cutoffs for significant liver fibrosis (>7kPa) and cirrhosis (>12.5kPa) were based on WHO 2024 CHB guidance. The WHO treatment criteria[262], and data collected against which treatment eligibility was determined are reported in **Table 6.1**. A

description of Fibroscan™ training events and feedback have been summarised in a report[267].

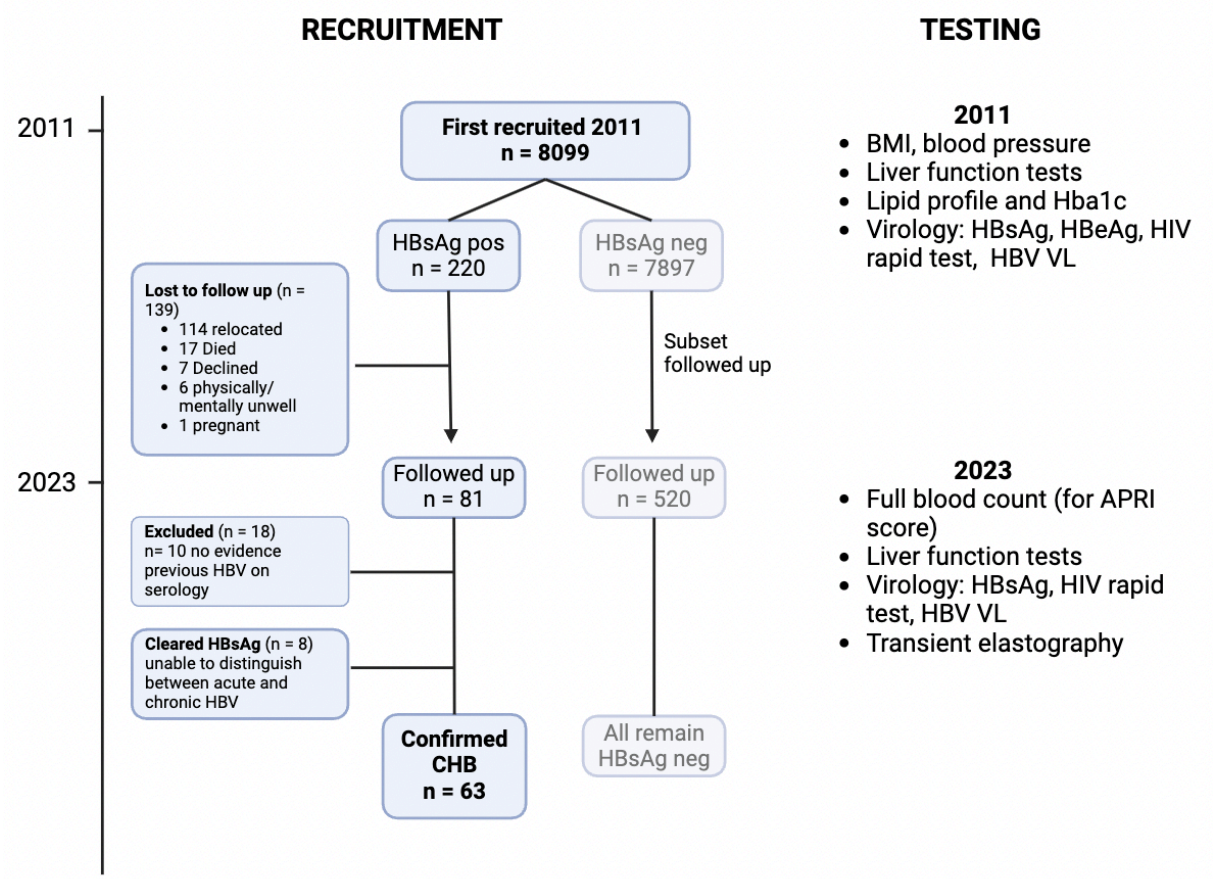


Figure 6.3. Flow diagram to show the timeline of recruitment, testing and follow up of Uganda liver disease (ULiDS) HBV cohort between 2011 - 2023 in Kalungu District, Uganda. HBV - Hepatitis B virus; CHB - chronic HBV infection; HBsAg - Hepatitis B surface antigen; APRI - aspartate aminotransferase to platelet ratio index; HIV - human immunodeficiency virus; VL - viral load

6.4.3 Results

Baseline characteristics and risk factors for liver disease

In 2011, 220/8099 (2.7%) individuals tested positive for HBsAg (**Figure 6.3**). Baseline characteristics for the whole cohort and individuals with HBV infection are presented in **Table 6.2.** The median age of the HBsAg positive cohort was 31 years (IQR 19 to 42), 44.5% were

female, 11.3% were living with HIV infection. Self-reported alcohol intake was recorded for the whole cohort in 2011 where 2919 (36.0%) reported ever drinking. In 2023, 28 (44.4%) reported drinking, with daily alcohol intake reported by 5/63 (7.9%) and weekly by 14/63 (22.2%).

Table 6.2. Uganda Liver Disease Study (ULiDS) cohort characteristics at baseline (2011) and in those followed up (2023)

Variable	2011 cohort (n = 8099)	HBsAg pos 2011			Confirmed CHB 2023 (n = 63)
		All (n = 220)	Followed up in CHB cohort in 2023 (n = 63)	Not followed up in CHB cohort in 2023 (n= 157)	
Sex					
Female Male	4557 (56.3%) 3452 (42.6%)	98 (44.5%) 122 (55.5%)	26 (41.3%) 37 (58.7%)	72 (45.9%) 85 (54.1%)	26 (41.3%) 37 (58.7%)
Age group					
13-19 20-29 30-39 40-49 50-59 >60	2481 (30.6%) 1508 (18.6%) 1349 (16.6%) 1095 (13.5%) 744 (9.2%) 922 (11.4%)	56 (25.5%) 51 (23.1%) 47 (21.4%) 31 (14.1%) 21 (9.5%) 14 (6.4%)	6 (9.5%) 12 (19.0%) 21 (33.3%) 13 (20.6%) 8 (12.7%) 3 (4.8%)	50 (31.8%) 39 (24.8%) 26 (16.5%) 18 (11.5%) 13 (8.2%) 11 (7.0%)	0 (0.0%) 5 (7.9%) 10 (15.9%) 23 (36.5%) 10 (15.9%) 15 (23.8%)
Infection status					
HBV positive HIV positive HCV positive Malaria positive	220 (2.7%) 608 (7.5%) 13 (0.2%) Not tested	220 (100%) 24 (11.3%) 0 (0.0%) Not tested	4 (6.3%)	20 (12.7%)	63 (100%) 5 (7.9%) Not repeated 2 (3.2%)
Frequency of alcohol consumption					
Never drinker Drinker Daily Weekly Monthly Yearly	5180 (64.0%) 2919 (36.0%)				35 (55.5%) 28 (44.4%) 5 (7.9%) 14 (22.2%) 3 (4.8%) 6 (9.5%)
Body mass index (BMI)					
Underweight (<18.5) Normal (18.5 – 24.99) Overweight (>25) Unknown	1772 (21.8%) 5095 (62.9%) 960 (11.9%) 272 (33.6%)	40 (18.1%) 143 (65.0%) 21 (9.5%) 16 (72.7%)	9 (14.2%) 42 (66.7%) 7 (11.1%) 5 (7.9%)	31 (19.7%) 101 (54.3%) 14 (8.9%) 11 (7.0%)	11 (17.4%) 41 (65.1%) 11 (17.5%)

*Variables only measured in 2011; ** As defined by MASLD criteria[268]; HBsAg - Hepatitis B surface antigen; CHB - chronic HBV infection; HIV - human immunodeficiency virus; HCV - Hepatitis C virus; HDL - high density lipoprotein; M - male; F - female.

Characteristics at follow up in 2023

In 2023, 81/220 individuals with positive HBsAg were followed up (139/220 were lost to follow up, reasons detailed in **Figure 6.3**). We performed quantitative HBsAg, antiHBc, antiHBs and HBV VL on 81 individuals. Using these results, I retrospectively identified eleven individuals with false positive HBsAg results from 2011 (HBsAg at the limit of detection in 2011, negative HBsAg, antiHBc and antiHBs in 2023) and excluded these from further analysis. In addition I identified eight individuals who were HBsAg negative and antiHBc positive in 2023. I could not distinguish between acute and chronic HBV infection and these individuals were therefore excluded from further analysis. Characteristics of the remaining 62 individuals with confirmed CHB at follow up are detailed in **Table 6.2**.

Eligibility for HBV treatment

Using the WHO 2024 HBV treatment guidelines[262] (**Table 6.1**), I determined treatment eligibility for 63 individuals with CHB who remained HBsAg positive in 2023. 24/63 (38%) individuals were eligible for treatment. **Figure 6.4** shows the number of individuals meeting each of the four main treatment criteria, including those who met more than one criterion. The majority of individuals (19/24, 79%) were eligible on the basis of having 'significant' fibrosis (either APRI >0.5 or elastography > 7kPa, no individuals had clinical evidence of cirrhosis). Only one individual met the threshold for treatment based on criterion 2 (HBV VL >2000 with ALT above ULN); they had a HBV VL of 7 $\log_{10}$ IU/ml in 2011 which had fallen to 353 IU/ml in 2023 without treatment and had no evidence of fibrosis using APRI or elastography. Five individuals were eligible for treatment due to HIV coinfection ('treat 3'), one of whom also met 'treat 1' fibrosis criteria. Only 1 individual met the 'treat 4' criteria of two ALT above ULN.

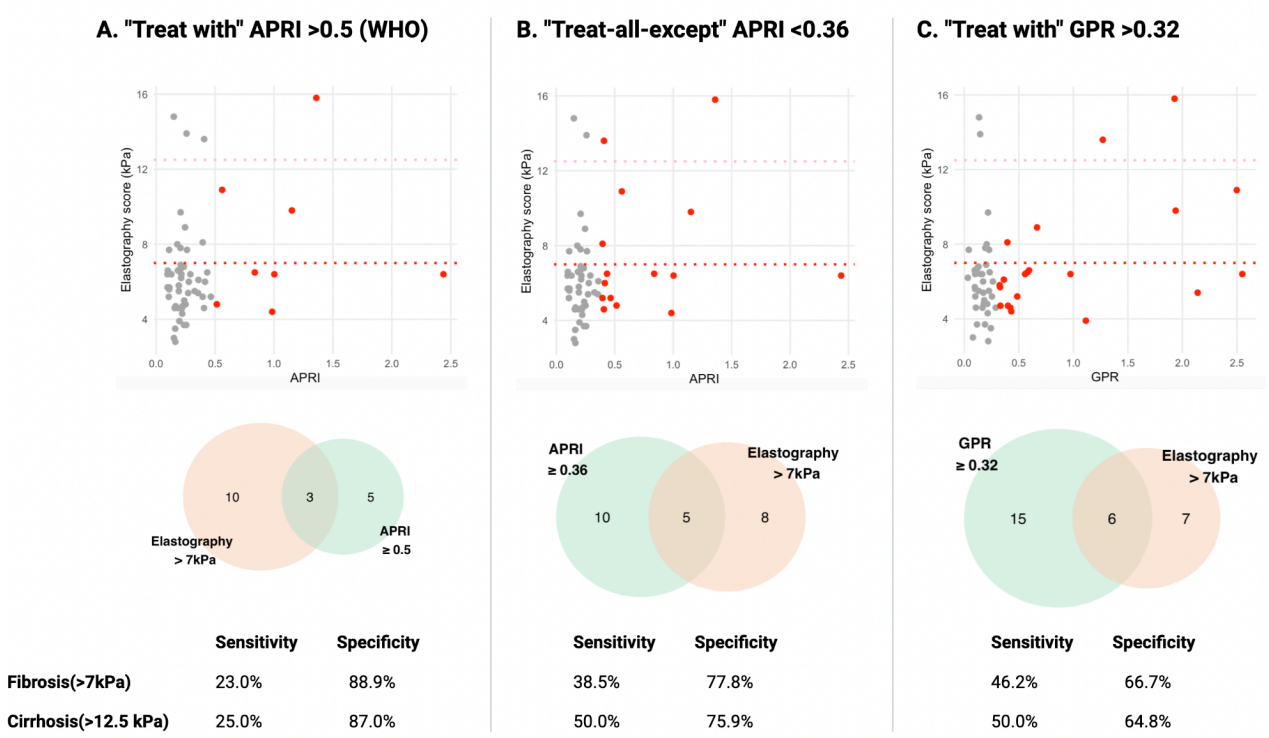
Given TE is not available in many clinical settings, I also determined treatment eligibility in a hypothetical scenario where TE was not available in this cohort. In this scenario only 14/62 (23%) individuals were eligible for treatment.



Figure 6.4. Eligibility for HBV treatment in Uganda liver disease study (ULiDS) HBV cohort based on WHO 2024 treatment guidelines[262].

Lack of concordance between non-invasive measures of liver disease in treatment-naive HBV mono-infected individuals

Using the WHO threshold of TE >7 kPa, 13/58 HBV mono-infected individuals were eligible for treatment. I investigated the sensitivity of different NIT for diagnosing fibrosis/cirrhosis in untreated individuals with CHB, using TE as the gold standard (**Figure 6.5**). I excluded HBV/HIV co-infected individuals (n = 5) for this analysis as they were already receiving HBV active antiviral therapy.



	Sensitivity	Specificity
Fibrosis(>7kPa)	23.0%	88.9%
Cirrhosis(>12.5 kPa)	25.0%	87.0%

	Sensitivity	Specificity
	38.5%	77.8%
	50.0%	75.9%

	Sensitivity	Specificity
	46.2%	66.7%
	50.0%	64.8%

Figure 6.5. Predictive value of routinely available biomarkers for detecting fibrosis or cirrhosis in treatment-naive individuals with CHB mono-infection in the Uganda liver disease (ULiDS) 2023, using TE as the gold standard. The graphs show the correlation between blood biomarker scores and TE. The red horizontal dotted lines show a TE of 7kPa (F2 fibrosis above which individuals are offered treatment), the pink horizontal dotted lines show a TE of 12.5kPa (F4 cirrhosis). Red dots highlight individuals who would be offered treatment based on blood biomarker alone. Venn diagrams show the number of individuals meeting who met one or both of the treatment eligibility criteria. Sensitivity and specificity of the blood biomarkers for predicting TE >7kPa and TE >12.5kPa are calculated using TE as the gold standard.

WHO 2024: "treat-with" APRI >0.5

Using a cut-off of APRI >0.5, 8/58 individuals were eligible for treatment. APRI identified a slightly different set of individuals for treatment than TE, only 3 individuals had agreement in eligibility between APRI and TE criteria (**Figure 6.5 A**). In this cohort, APRI >0.5 had a low sensitivity for the prediction of fibrosis and cirrhosis; an APRI threshold of >0.5 had a

sensitivity of only 23.0% for predicting a TE score >7 kPa (F2 fibrosis) and 25.0% for predicting a TE score >12.5 kPa (F4 cirrhosis).

HEPSANET: "treat-all-except" APRI <0.36

The 'treat-all-except' paradigm utilises a more inclusive approach to treatment, where an optimised rule-out threshold of APRI <0.36 is used to exclude "low risk" individuals not requiring treatment, i.e. all individuals with APRI > 0.36 are offered treatment. Using this lower APRI cut-off, 15/58 individuals were eligible for treatment, with an increased sensitivity of 38.5% for detection an elastography score >7 kPa (F2 fibrosis) and 50.0% for predicting an elastography score >12.5 kPa (F4 cirrhosis) **(Figure 6.5 B)**.

PROLIFICA: "treat-with" GPR >0.32

Using a GPR >0.32, 21/58 individuals were eligible for treatment. Sensitivity was 46.2% for predicting fibrosis (>7kPa) and 50.0% for cirrhosis (>12.5 kPa) (**Figure 6.5 C**). GPR >0.32 was the only criteria to identify the single individual meeting the "treat 2" criteria of HBV VL >2000 IU/ml and ALT >ULN (GPR 0.324, APRI 0.19). For this analysis I used the GPR cut-off determined by Lemoine et al[265] based on their cohort in Gambia. However in this cohort, the GPR threshold that maximises the sum of the sensitivity and specificity (the metric used to determine the Gambia cut-off) is higher at 0.60.

"Treat-all" approach

If a 'treat-all' approach was used in this population, 39/63 individuals who would have been ineligible using the WHO 2024 treatment criteria would also be offered treatment.

Which individuals are missed with blood biomarker based NIT?

In this cohort, ten individuals with TE > 7kPa who should have been eligible for treatment would have been missed in a setting using an APRI >0.5 alone. Only one of these individuals met one of the other treatment criteria, 9/10 (90%) would have not received treatment. Two individuals with cirrhosis (TE >12.5 kPa) would not have met the threshold for treatment by any of the non-invasive biomarkers scores/cut-offs or been captured by other WHO treatment criteria. One individual had a HBV VL of ~90,000 IU/ml, the other ~900 IU/ml, however both individuals had platelet count, ALT, AST and GGT within the reference range. Neither had co-infection with HIV, HCV or malaria and neither reported drinking alcohol. One individual had a BMI of 25.3 in 2011 (overweight) but had fallen to the normal range in 2023.

HBV VL as indication for treatment

Fifteen individuals had VL measured in both 2011 and 2023, 56 individuals in 2023 alone and 42 individuals in 2011 alone. The distribution of VL in HBV mono-infected untreated individuals is shown in **Figure 6.6**, a bimodal distribution is seen in **Figure 6.6 A** as has been described previously[108]. Two individuals (2/14, 14.3%) had an HBV VL <2000 IU/ml in 2011 and >2000 IU/ml in 2023). All individuals with HIV coinfection on HBV-active ART had undetectable HBV VL in 2023, no other individuals were on antiviral treatment. Only one individual met the criteria for treatment based on HBV VL/ALT in 2011 or 2023 (HBV VL > 2000 IU/ml and an ALT >ULN).

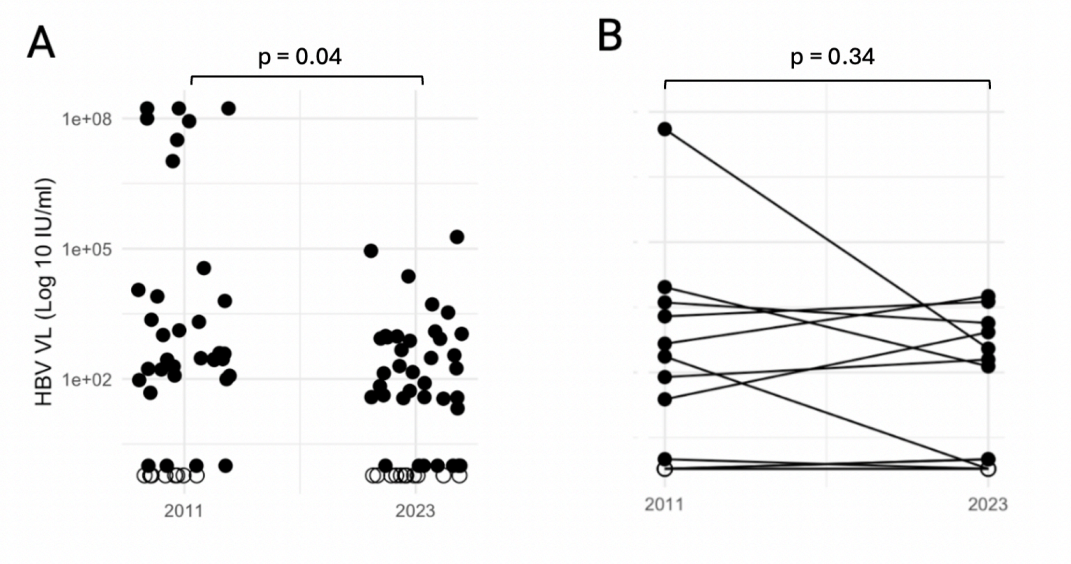


Figure 6.6. HBV VL in 2011 and 2023 in HBV mono-infected untreated individuals from the Uganda liver disease study (ULiDS). A) Individuals with HBV VL testing performed at a single time point (p = 0.04, Mann-Witney U test). B) Individuals with HBV VL testing performed at two time points (p = 0.34, paired t-test).
Where an individual had samples at both time points, these are joined by a black line. Where HBV VL was undetected, the value is plotted as an empty circle (o), where HBV VL was below the limit of quantification (<20 IU/ml), the value is plotted as 1.

6.4.4 Conclusion

I applied the 2024 WHO HBV treatment criteria to a cohort of individuals with CHB in rural Uganda, establishing the proportion of individuals eligible for treatment, to guide local treatment strategy. I highlight how availability of diagnostics impacts on ability to identify individuals who might benefit from treatment, in particular the pitfalls in using NIT scores to predict fibrosis/cirrhosis. I evaluate alternative approaches to risk stratification and treatment.

When implementing the new WHO guidelines in settings without reliable access to TE or HBV VL, NITs such as APRI become attractive for assessing treatment eligibility based on liver disease. In this population, 24/63 (38%) were eligible for treatment using all data

available (**Table 6.1**), this fell to 14/63 (23%) in a hypothetical scenario in which TE was not available and APRI >0.5 was relied on to identify those with liver fibrosis and cirrhosis. My findings raise concerns about the adequacy of the APRI threshold >0.5 as a predictor of liver fibrosis and cirrhosis. In this population, use of the APRI >0.5 threshold fails to identify the majority of individuals with fibrosis and cirrhosis (sensitivity 23% and 25% respectively) and leads to missed opportunities for treatment. The sensitivity I observe contrasts to that referenced in the WHO HBV guidelines, in which an APRI >0.5 was associated with a sensitivity of 71.7% (95% CI 67.1–75.8%) for identifying ≥F2 among people with CHB infection. The review acknowledges a lack of evidence from the WHO African region[269]. The possibility of false negative results is raised, and mitigated against by saying that people with CHB infection may meet other eligibility criteria for antiviral therapy (such as elevated HBV VL), however only 10% of individuals with a false negative APRI score met another treatment criteria in this cohort. I demonstrate improved, but still limited, sensitivity using a 'treat-all-except' threshold of APRI <0.36 or a 'treat-with' threshold of GPR >0.32, the increase in sensitivity is coupled with a fall in specificity.

In this cohort, a combination of ALT/HBV VL (‘treat 2’) and a persistently positive ALT (‘treat 4’) identified one individual each for treatment (1.6%). This contrasts with the WHO estimates that ALT/DNA would capture an estimated 20-35% of all HBsAg positive people and that persistently positive ALT estimated at 20%. Low eligibility in this cohort could be driven by the older age at follow up, typically associated with lower HBV VL, in addition to the low frequency of ALT and HBV VL testing compared to the recommended 12 monthly tests; ALT levels fluctuate and require longitudinal monitoring to determine the trend. However despite these caveats it raises the question about the incremental benefit of performing HBV VL in this population for the purposes of treatment stratification, especially given that HBV VL

testing is not readily available in many settings, and where available, it is expensive either to the government or the individual paying out-of-pocket.

Together, the lack of access and expense of TE the gold standard NIT, the poor sensitivity of cheaper, accessible NITs such as APRI and the lack of access and expense of HBV VL raise some fundamental questions. Is it preferable to risk over-diagnosis or under-diagnosis of liver disease in this setting? While over-diagnosis might increase unnecessary treatment and increase drug costs, under-diagnosis risks failing to identify individuals who require timely intervention to prevent further liver disease related morbidity and mortality. In the absence of appropriate risk stratification tools, should we be treating all? A treat-all approach would effectively reduce HBV-associated morbidity and mortality, reduce the cost of evaluation, overcome challenges with shifting eligibility over time, and negate the issue that different NITs identify different sets of individuals for treatment. Furthermore it would provide a more effective suppression of the HBV reservoir by reducing high VLs in individuals not otherwise eligible for treatment (a benefit not included in current cost estimates[270]) therefore reducing horizontal and mother to child transmission by protecting all pregnancies in addition to acting as HIV PrEP (if TFV used). However acceptability, the burden of life-long treatment and economic feasibility remains a concern[270].

The small sample size and high loss to follow-up, particularly due to migration of younger individuals, presents challenges for longitudinal assessment of treatment eligibility and health outcomes. Although TE is accepted globally as an acceptable alternative to liver biopsy for the identification of liver fibrosis and cirrhosis in CHB, TE itself (as with all tests) has limitations in sensitivity and specificity (sensitivity using a cut-off of >6.0 to 8.0 kPa for F2 fibrosis 75.1% (95% CI 72.2–77.7%), specificity 79.3% (95% CI 76.2–82.2%) compared to liver biopsy which itself has issues with inter-reporter reliability and sampling error[269]). Individuals were asked to fast prior to testing in this cohort, however false positive TE scores

may occur if individuals were not fasting. People with hepatitis flares or acute hepatitis will have falsely high liver stiffness measurements however no individuals with high TE and low APRI had raised liver ALT or AST. In a clinical setting a repeat TE may be performed to clarify the clinical picture however this was not possible here. I planned to evaluate the TREAT-B score (based on ALT and HBeAg status) however HBeAg testing was not available in 2023, highlighting issues with assay availability even in well resourced research settings.

6.5. Application of HBV WGS in four settings in the WHO African region

6.5.1 Introduction

There are limited numbers of publicly available HBV sequences from many countries in the WHO African region, particularly Uganda (1 available in Genbank) and DRC (1 available), with more available from Kenya (25) and South Africa (70). Genotypes A, D and E predominate in these regions, with evidence of some recombinants (predominantly A/E or D/E) and rare reports of genotypes B and C in South Africa[47]. There is limited evidence to document the prevalence of transmitted RAMs in the WHO African region, and current estimates vary widely (from <1% to 37%) by setting and therapeutic exposure[104]. Globally, rates of acquired drug resistance are low for both agents in NA naive individuals, however baseline sequencing is not routinely performed, making the distinction between acquired and transmitted drug resistance challenging. Furthermore, there is evidence across multiple settings of individuals with detectable viraemia on TFV despite good adherence to therapy[271–273], raising the question of whether novel TFV RAMs, which are not included in available clinical resistance detection tools, are influencing treatment response.

Here I summarise the sequencing outputs from the ULiDS cohort described above, in addition to preliminary data from Kenya, SA and the DRC. I explore the distribution of genotypes, prevalence of known RAMs and VEMs, and aim to identify novel RT polymorphisms in individuals with virological breakthrough (VBT) during TFV-based therapy. Finally I use WGS to investigate transmission patterns in household clusters across different settings (**Figure 6.7**).

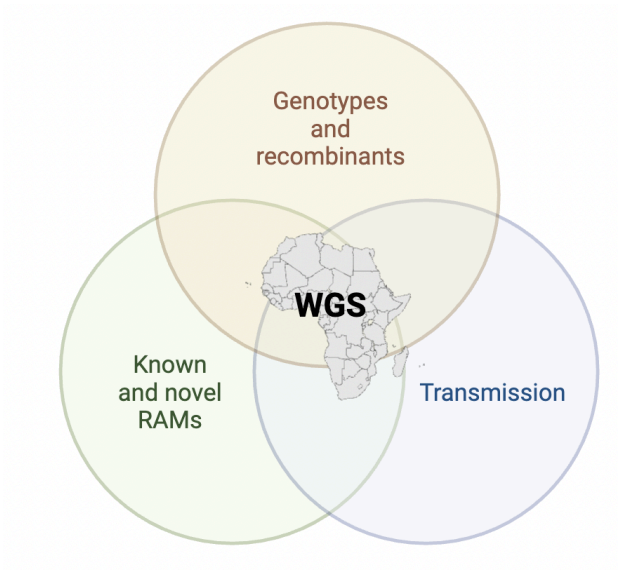


Figure 6.7. Application of WGS to four settings in WHO African region leads to shared insights into genotypes, recombinants, known and novel resistance associated mutations (RAMs) and transmission.

6.5.2 Methods

Settings

I have described the settings in Uganda, Kenya, SA and DRC and respective samples in detail in **Chapter 2**, **Table 2.1** and **Appendix 2** (**Supplementary Tables 6-9**). I performed HBV WGS on samples from SA and DRC in Oxford, and travelled to Uganda in May 2024 and Kenya in October 2024 to perform sequencing in-country, whilst additionally supporting building of local capacity through training.

HBV library preparation, sequencing and bioinformatic analysis

I used the HEP-TILE protocol for Nanopore that I developed in **Chapter 4**[146,274]. Key variations in primer scheme (determined by time point in project at which sequencing was performed) and platform are summarised in **Table 6.3**. All libraries were sequenced on

R10.4.1 flow cells. Consensus HBV sequences were generated by Chris Kent (University of Birmingham) using hbv-fieldbioinformatics[150] and I analysed consensus sequences as described in **Chapter 2, section 2.2.3**.

Table 6.3. Details of the Hepatitis B virus whole genome sequencing protocol used for different cohorts. Primer sequences available online[275]

	Democratic Republic of the Congo	South Africa	Uganda	Kenya
Primer scheme	hbv/500/v1.1.0	hbv/600/v2.1.0	hbv/600/v2.1.0	hbv/600/v2.1.0
Platform	MinION	MinION	MinION	GridION

6.5.3 Results

Overview of sequences generated

I sequenced a total of 143 samples, I obtained 81 full genomes (>98% coverage), 26 near-full genomes (80-98%) and 31 partial genomes (**Figure 6.8, Table 6.4**); genome coverage varied by HBV VL.

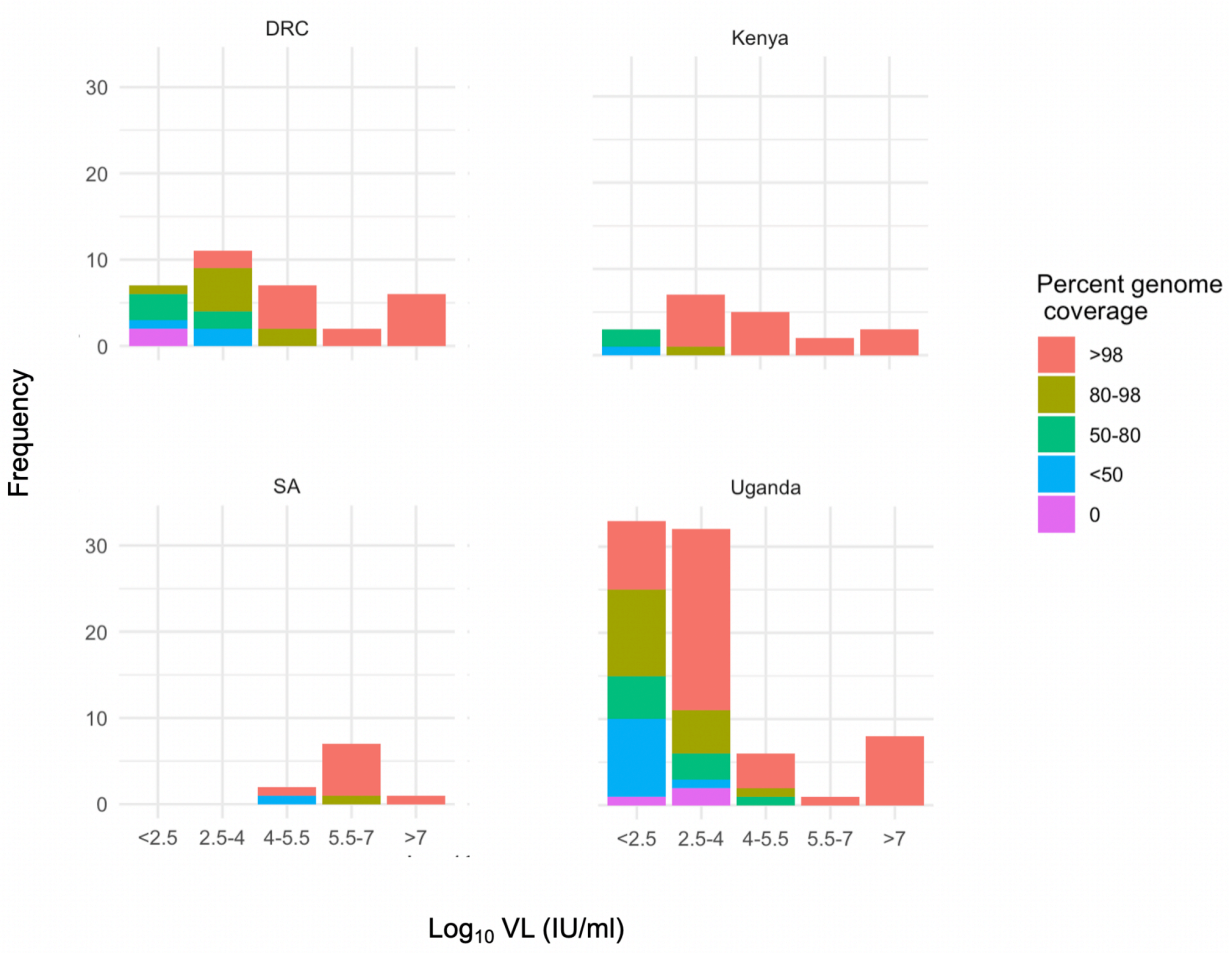


Figure 6.8. Percent HBV genome coverage split by HBV VL in four settings in the WHO African region. X axis shows VL (log_{10} IU/ml), y axis shows the frequency of sequences within each genome coverage category. DRC samples were enriched using hbv/500/v1.1.0 primers (amplicon 4 drop out at low VL), samples from other cohorts were enriched using hbv/600/v2.1.0 primers.
DRC - Democratic Republic of the Congo; SA - South Africa

Table 6.4. Comparison of HBV WGS coverage across four settings in the WHO African region.

Percent genome coverage	Democratic Republic of the Congo (n=33)	South Africa (n=10)	Uganda (n=80)	Kenya (n=20)
> 98%	15/33 (45.5%)	8/10 (80.0%)	42/80 (52.5%)	16/20 (80.0%)
80 - 98%	8/33 (24.2%)	1/10 (10.0%)	16/80 (20.0%)	1/20 (5.0%)
50 - 80%	5/33 (15.1%)	0/10 (0.0%)	9/80 (11.3%)	2/20 (10.0%)
1 - 50%	3/33 (9.1%)	1/10 (10.0%)	10/80 (12.5%)	1/20 (5.0%)
0%	2/33 (6.1%)	0/10 (0.0%)	3/80 (3.8%)	0/20 (0.0%)

The distribution of genotypes and recombinants differed between settings. Sequences from Kenya were all genotype A (subgenotype A1) and those from SA were predominantly genotype A (subgenotype A1) with one genotype C sequence (subgenotype C2) which has been identified at low prevalence in SA previously[47]. DRC and Uganda had a more diverse mix of sequences. DRC sequences were genotype A (subgenotypes A1, A3, A6), genotype D (subgenotype D3) and D/E recombinants (previously only genotypes A and E identified in this country). Uganda sequences were genotype A (subgenotype A1) and A/D, A/D/E and D/E recombinants (**Table 6.5**). D/E recombinants have been observed in a previous study in Uganda that used partial genome sequencing of a Pol fragment[52]. Maximum likelihood trees demonstrating a range of genotypes from Uganda (**Figure 6.9**) and South Africa (**Figure 6.10**) are shown below.

Table 6.5. Distribution of HBV genotypes across four settings in the WHO African region.

Genotype	DRC (n=33)	South Africa (n = 10)	Uganda (n = 80)	Kenya (n = 20)
A	12/33 (36.4%)	8/10 (80.0%)	57/80 (71.3%)	19/20 (95.0%)
C	-	1/10 (10.0%)	-	-
D	1/33 (3.0%)	-	-	-
E	12/33 (36.4%)	-	-	-
A/D	-	-	2/80 (2.5%)	-
A/D/E	-	-	3/80 (3.8%)	-
D/E	3/33 (9.1%)	-	5/80 (6.3%)	-
Not determined*	5/33 (15.2%)	1/10 (10.0%)	13/80 (16.3%)	1/20 (5.0%)

*Genotype not determined for sequences with <50% genome coverage
DRC - Democratic Republic of the Congo

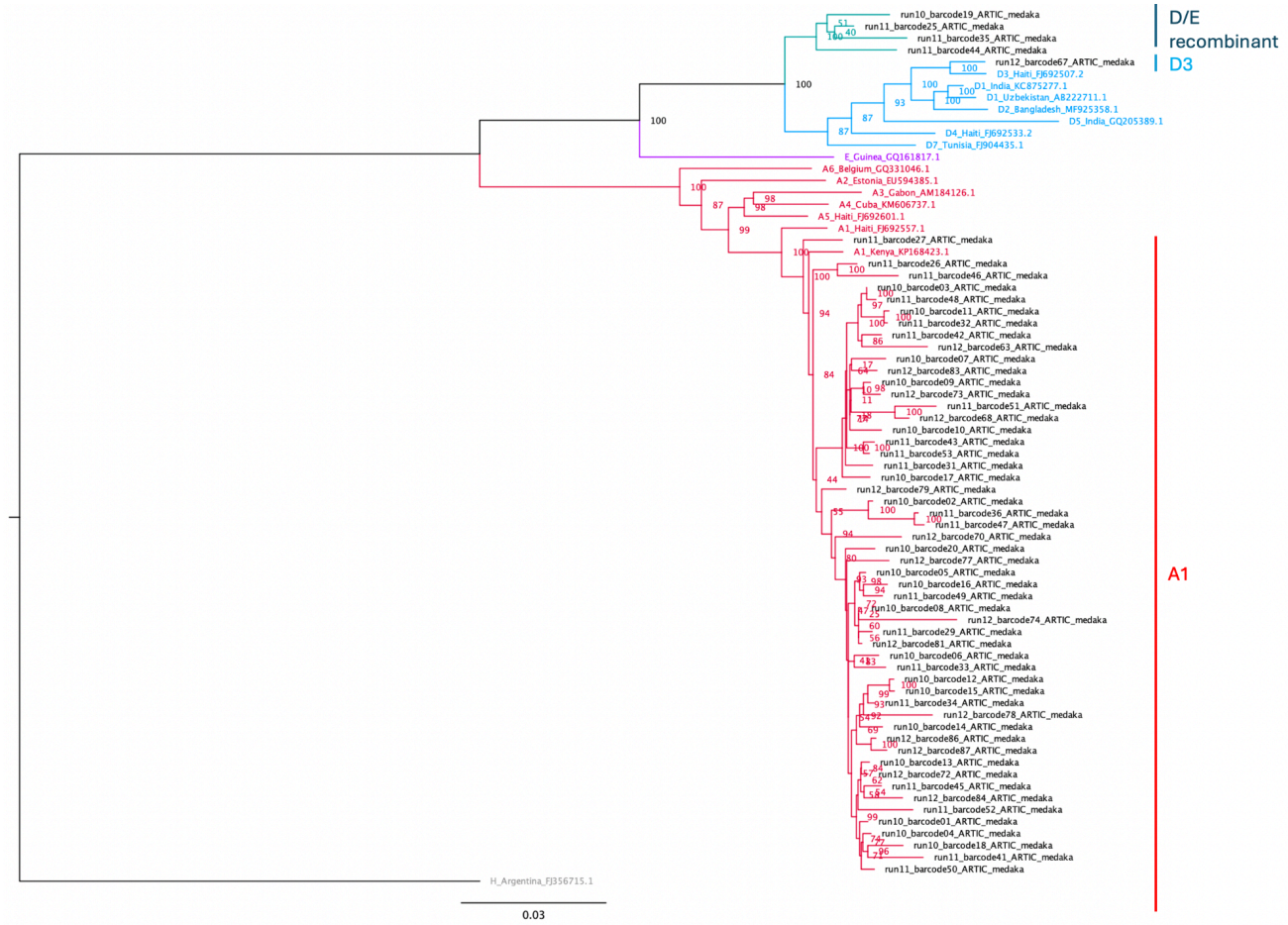


Figure 6.9. Maximum likelihood phylogenetic tree of HBV WGS from Uganda with genotype A, D and E reference strains[53]**.** Constructed with sequences with >80% genome coverage. Reference strain names are coloured, sample names are black. Bootstrap values displayed to the right of nodes. The lengths of the horizontal bars indicate the number of nucleotide substitutions per site.

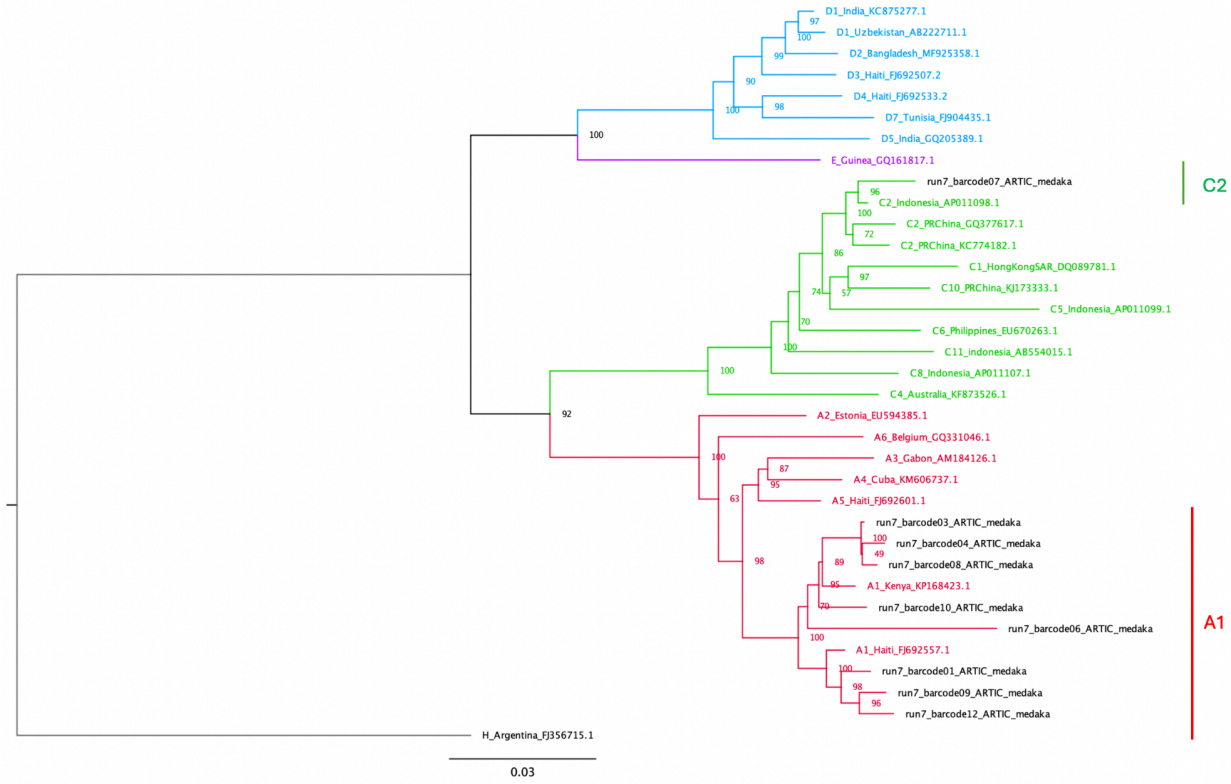


Figure 6.10. Maximum likelihood phylogenetic tree of HBV WGS from South Africa with genotype A, C, D and E reference strains[53]**.** Constructed with sequences with >80% genome coverage. Reference strain names are coloured, sample names are black. Bootstrap values displayed to the right of nodes. The lengths of the horizontal bars indicate the number of nucleotide substitutions per site.

RAMs in NA-naive and NA-experienced individuals

I analysed HBV sequences split by individuals' treatment status; the full RT region was available for 110 sequences from NA-naive individuals and 6 NA-experienced individuals (all treated with TFV). I determined the frequency of polymorphisms at positions of 10 established RAMs, and 7 putative TFV RAMs (see **Chapter 2**, **Table 2.7**). I have shown the wild type reference amino acid for all genotypes and subgenotypes for each position of interest in **Table 6.6**. Some putative TFV RAMs are polymorphisms in the subgenotype in which they were originally identified (e.g rt126Y and rt269I from the "CYEI" combination in gt C2, **Table 6.6**), however these amino acids are wildtype in other genotypes.

In NA-naive individuals (n=110), 2/110 (1.8%) sequences carried the rt180M/rt204V combination of mutations, associated with reduced sensitivity to ETV and 3TC/LdT resistance (**Figure 6.11**). Both individuals with rt180M/rt204V mutations were from the Uganda setting (2/64, 3.1%), had genotype A infections and were co-infected with HIV/HBV. I identified putative TFV RAMs in some sequences (rt126Y in 67.3%, rt153W in 80.0% and rt269L in 8.2% of sequences), however I did not observe any RAM combinations that confer clinical resistance (**Table 6.7**).

In NA-experienced individuals (n=6), no sequences contained established RAMs (**Figure 6.11**). Of the putative TFV RAMs, I identified rt126Y (66.7%) and rt153W (83.3%). All 6 NA-experienced individuals were infected with a subgenotype A1 strain, both rt126Y and rt153W are wildtype residues in the subgenotype A1 reference strains. No individuals had combinations of RAMs that have previously been associated with TFV resistance.

Table 6.6. Amino acid residues found at established and putative RAM sites in HBV reference strains. Red = amino acid associated with drug resistance, Green = amino acid not associated with drug resistance. Reference strains with Genbank IDs from McNaughton et al.[157]

Geno	Subgeno	Established RAMs										Putative tenofovir RAMs						
		rt80	rt169	rt173	rt180	rt181	rt184	rt202	rt204	rt236	rt250	rt106	rt126	rt134	rt153	rt194	rt204	rt269
A	A1(1) KP168423	L	I	V	L	A	T	S	M	N	M	S	Y	D	W	A	M	I
	A1(2) FJ692557	L	I	V	L	A	T	S	M	N	M	S	Y	D	W	A	M	I
	A2 EU594385	L	I	V	L	A	T	S	M	N	M	S	Y	D	W	A	M	I
	A3 AM184126	L	I	V	L	A	T	S	M	N	M	S	Y	D	R	A	M	I
	A4 KM606737	L	I	V	L	A	T	S	M	N	M	S	H	D	R	A	M	I
	A5 FJ692601	L	I	V	L	A	T	S	M	N	M	S	Y	D	R	A	M	I
	A6 GQ331046	L	I	V	L	A	T	S	M	N	M	S	Y	D	W	A	M	I
B	B1 D23679	L	I	V	L	A	T	S	M	N	M	S	H	N	W	A	M	I
	B2(1) JQ801514	L	I	V	L	A	T	S	M	N	M	S	H	N	R	A	M	I
	B2(2) GU815637	L	I	V	L	A	T	S	M	N	M	S	H	D	R	A	M	I
	B3 AP011085	L	I	V	L	A	T	S	M	N	M	S	H	D	R	A	M	I
	B4 AB073835	L	I	V	L	A	T	S	M	N	M	S	H	D	R	A	M	I
	B6 AB287314	L	I	V	L	A	T	S	M	N	M	S	H	D	R	A	M	I
C	C1 DQ089781	L	I	V	L	A	T	S	M	N	M	S	H	D	R	A	M	I
	C2(1) KC774182	L	I	V	L	A	T	S	M	N	M	S	H	D	R	A	M	L
	C2(2) GQ377617	L	I	V	L	A	T	S	M	N	M	S	H	D	R	A	M	I
	C2(3) AP011098	L	I	V	L	A	T	S	M	N	M	S	H	D	R	A	M	L
	C4 KF873526	L	I	V	L	A	T	S	M	N	M	S	H	D	R	A	M	I
	C5 AP011099	L	I	V	L	A	T	S	M	N	M	S	H	D	R	A	M	I
	C6 EU670263	L	I	V	L	A	T	S	M	N	M	S	Y	D	R	A	M	I
	C8 AP011107	L	I	V	L	A	T	S	M	N	M	S	Y	D	R	A	M	I
	C10 KJ173333	L	I	V	L	A	T	S	M	N	M	S	H	D	R	A	M	I
	C11 AB554015	L	I	V	L	A	T	S	M	N	M	S	H	N	R	A	M	I
	C(UA1) KC774298	L	I	V	L	A	T	S	M	N	M	S	H	D	R	A	M	I
	C(UA2) DQ089802	L	I	V	L	A	T	S	M	N	M	C	Y	E	R	A	M	I
D	D1(1) AB222711	L	I	V	L	A	T	S	M	N	M	S	H	D	R	A	M	I
	D1(2) KC875277	L	I	V	L	A	T	S	M	N	M	S	H	D	R	A	M	I
	D2 MF925358	L	I	V	L	A	T	S	M	N	M	S	R	D	R	A	M	I
	D3 FJ692507	L	I	V	L	A	T	S	M	N	M	S	H	D	R	A	M	I
	D4 FJ692533	L	I	V	L	A	T	S	M	N	M	S	H	D	W	A	M	I
	D5 GQ205389	L	I	V	L	A	T	S	M	N	M	S	H	D	R	A	M	L
	D7 FJ904435	L	I	V	L	A	T	S	M	N	M	S	H	D	W	A	M	I
E	E GQ161817	L	I	V	L	A	T	S	M	N	M	S	Y	D	R	A	M	I
F	F1 HM585194	L	I	V	L	A	T	S	M	N	M	S	H	N	R	A	M	I
	F2 DQ899143	L	I	V	L	A	T	S	M	N	M	S	H	N	R	A	M	I
	F3 MH051986	L	I	V	L	A	T	S	M	N	M	S	H	N	R	A	M	I
	F4 KJ843175	L	I	V	L	A	T	S	M	N	M	S	H	N	R	A	M	I
G	G AB056513	L	I	V	L	A	T	S	M	N	M	S	Y	D	R	A	M	I
H	H FJ356715	L	I	V	L	A	T	S	M	N	M	S	H	H	R	A	M	I
I	I1 AB562463	L	I	V	L	A	T	S	M	N	M	S	Y	D	Q	A	M	I
	I2 FJ023669	L	I	V	L	A	T	S	M	N	M	S	Y	D	R	A	M	I
J	J AB486012	L	I	V	L	A	T	S	M	N	M	S	H	D	R	A	M	I

Geno - genotype; Subgeno - subgenotype and Genbank ID; rt - reverse transcriptase; RAM - resistance associated mutation.

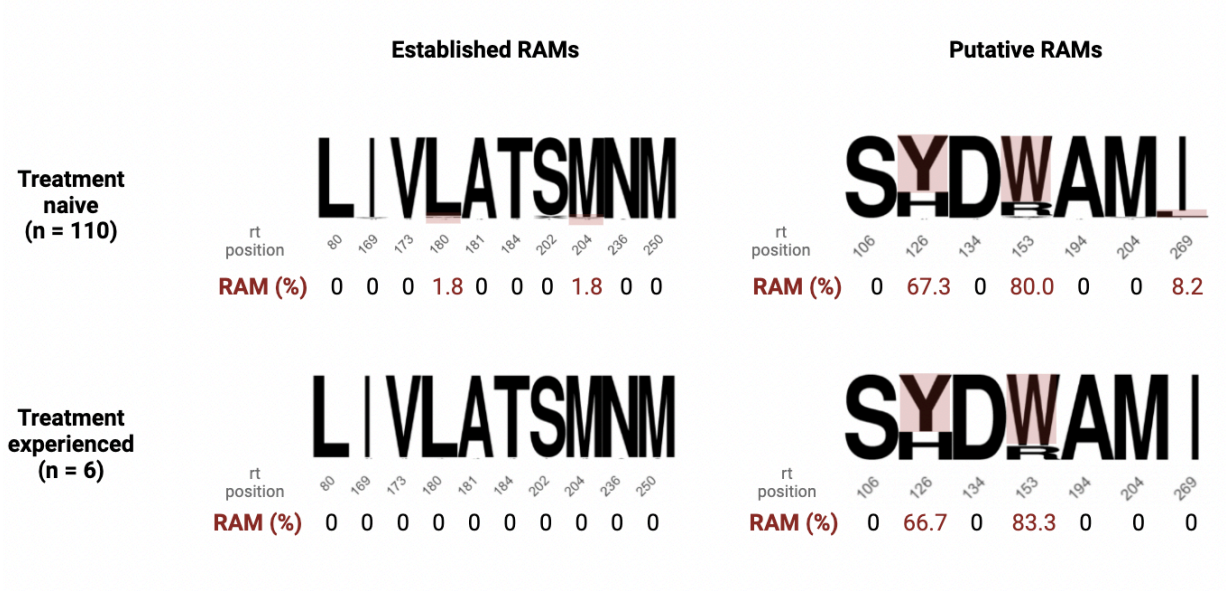


Figure 6.11. Presence of established and putative RAMs in 110 treatment naive and 6 treatment experienced individuals with CHB. The height of each letter represents the percentage of sequences containing that amino acid. Pink shading highlights amino acids associated with resistance, RAM (%) indicates the percentage of sequences containing the amino acid associated with resistance, RT numbering is based on the X02763 reference sequence. Amino acid representations produced using WebLogo, University of California, Berkeley[162].

Table 6.7. Prevalence of combinations of RAMs associated with reduced or intermediate susceptibility to NAs.

HBV variant	NA drug agent					NA naive (n = 110)				NA experienced
	3TC	LdT	ETV	ADV	TFV	Uganda (n = 64)	Kenya (n = 17)	DRC (n = 26)	SA (n = 3)	All settings (n = 6)
Established RAMs										
L180M + M204VI	R	R	I	I	S	2/64 (3.1%)	0	0	0	0
A181T/V	I	I	S	R	I	0	0	0	0	0
N236T	S	S	S	R	I	0	0	0	0	0
L180M + M204V/1 +/- I169T +/- V173L +/- M250V	R	R	R	S	S	0	0	0	0	0
L180M + M204V/I +/- T184G +/- S202I/G	R	R	R	S	S	0	0	0	0	0
Putative TFV RAMs										
rtS106C + rtH126Y + rtD134E + rtM204I/V + rtL269I [163] Known as “CYEI”	-	-	-	-	R	0	0	0	0	0
rtA194T + rtL180M + rtM204V [164]	-	-	-	-	R	0	0	0	0	0
rtR153W + rtI163V + rtL180M + rtM204V[165]	-	-	-	-	R	0	0	0	0	0

RAM - resistance associated mutation; NA - nucleos(t)ide analogue; 3TC - lamivudine; LdT - telbivudine; ETV - entecavir; ADV - adefovir; TFV - tenofovir; DRC - Democratic Republic of the Congo; SA - South Africa; R - resistant, I - intermediate/reduced susceptibility, S - sensitive.

As I did not identify combinations of RAMs which have previously been associated with TFV resistance in any of the 6 individuals with VBT on TFV, I searched for novel polymorphisms across the whole RT region. I identified polymorphisms at 19 other sites in RT (**Figure 6.12**). The combinations of RT polymorphisms found in each sequence is shown in **Table 6.8**. The polymorphism rt263E has previously been observed in a case report of an individual with genotype D infection with VBT on TFV[87,276]. Four of the other RT positions have previously been identified in individuals with VBT on TFV however with different amino acids from those which I observed here (122L, 163V, 217R, 333Q)[87]. Only rt125E was enriched in NA-experienced individuals compared to NA-naive individuals (Fisher's exact test, adjusted p <0.01), **Table 6.9**.

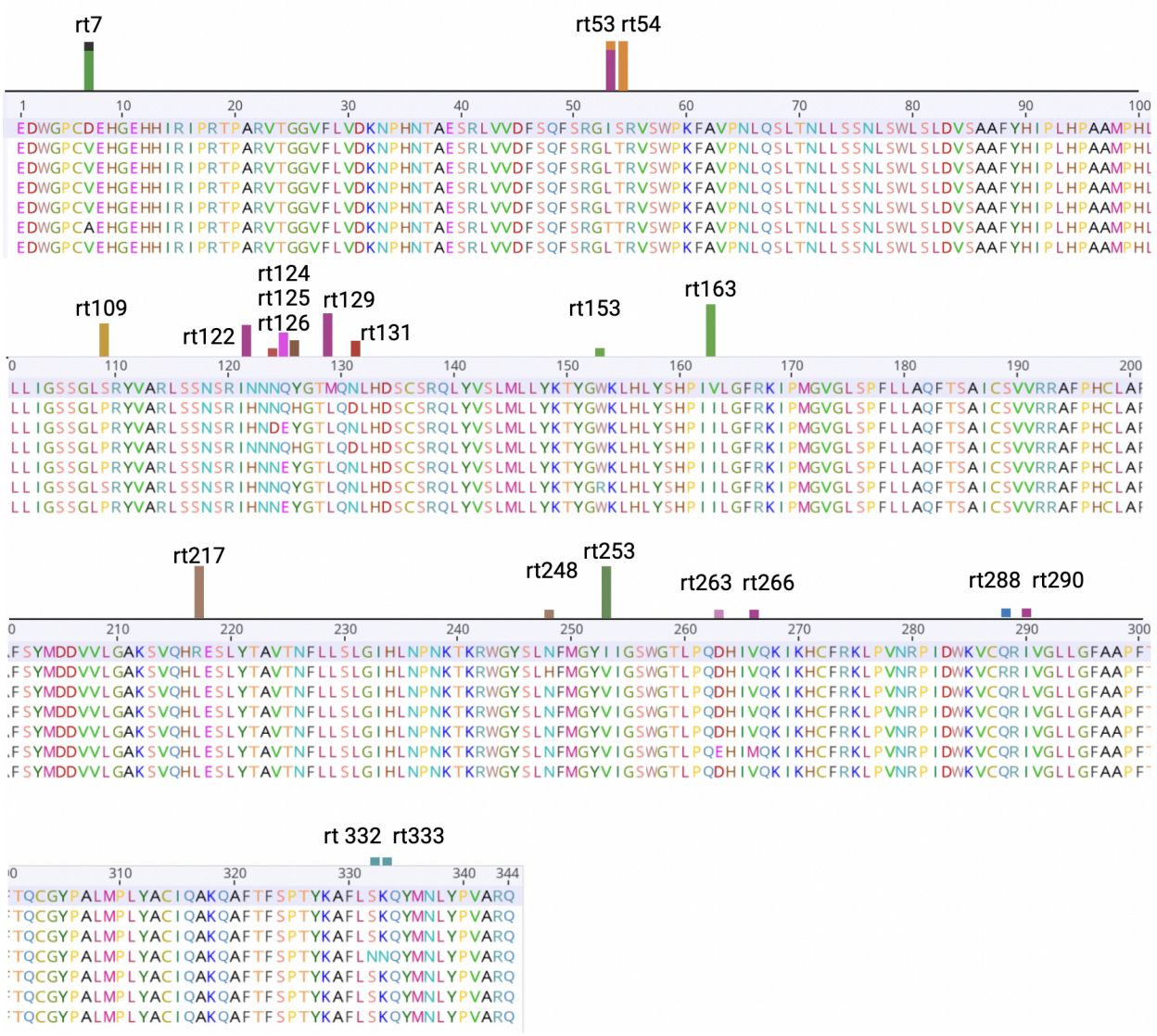


Figure 6.12. Polymorphic sites in HBV RT sequences from six individuals with virological breakthrough on tenofovir. The top sequence shaded in blue indicates the genotype A reference consensus sequence on which the numbering of rt amino acid position is based and from which polymorphisms are defined (ref strain X02763). Coloured vertical bars show sites where amino acid differs from the reference strain; 2 sites (rt126 and rt153) previously analysed in **Figure 6.11** and 19 novel sites. Height of the vertical bar indicates the number of sequences in which the polymorphism was identified, colour of the bar indicates the amino acid identified.

Table 6.8. RT polymorphisms in six individuals with VBT on tenofovir. List of all polymorphisms detected compared to ref strain X02763. Numbered by ref strain X02763.

Patient ID	Setting	Polymorphisms
SHBV-A3-026	Kenya	rt7V, rt53L, rt54T, rt126H, rt129L, rt131D, rt153W, rt163I, rt217L, rt253V
1191	SA	rt7V, rt53L, rt54T, rt124D, rt125E, rt129L, rt153W, rt163I, rt217L, rt253V, rt290L
SHBV-A3-042	Kenya	rt7V, rt53L, rt54T, rt126H, rt129L, rt131D, rt153W, rt163I, rt217L, rt253V, rt332N, rt333N
1175	SA	rt7V, rt53L, rt54T, rt125E, rt129L, rt153W, rt163I, rt217L, rt253V
1113	SA	rt7A, rt53T, rt129L, rt153R, rt163I, rt217L, rt253V, rt263E, rt266M
1143	SA	rt7V, rt53L, rt54T, rt125E, rt129L, rt153W, rt163I, rt217L, rt253V

Table 6.9. Frequency of HBV polymorphisms identified in consensus sequences derived by Nanopore sequencing of NA-experienced and NA-naive individuals with genotype A infection. Sites excluded if all sequences in both patient groups carried the same amino acid. RT numbering based on ref strain X02763.

Rt position	Amino acid	Treated n, %	Untreated n, %	P value	Adjusted p value
7	A	1/6 (16.7%)	3/74 (4.1%)	0.272	1.000
	V	5/6 (83.3%)	23/74 (31.1%)	0.018*	1.000
53	L	5/6 (83.3%)	64/74 (86.5%)	1.000	1.000
	T	1/6 (16.7%)	0/74 (0.0%)	0.075	1.000
109	P	5/6 (83.3%)	65/74 (87.8%)	0.564	1.000
	S	1/6 (16.7%)	8/74 (10.8%)	0.523	1.000
122	H	5/6 (83.3%)	64/74 (86.5%)	1.000	1.000
	N	1/6 (16.7%)	10/74 (13.5%)	1.000	1.000
124	D	1/6 (16.7%)	0/74 (0.0%)	0.075	1.000
	N	5/6 (83.3%)	69/74 (93.2%)	0.384	1.000
125	**E**	**3/6 (50.0%)**	**0/74 (0.0%)**	**<0.001***	**0.016 ****
	Q	**3/6 (50.0%)**	**74/74 (100.0%)**	**<0.001***	**0.016 ****
126	H	2/6 (33.3%)	24/74 (32.4%)	1.000	1.000
	Y	4/6 (66.7%)	49/74 (66.2%)	1.000	1.000
131	D	2/6 (33.3%)	18/74 (24.3%)	0.637	1.000
	N	4/6 (66.7%)	55/74 (74.3%)	0.650	1.000
153	R	1/6 (16.7%)	2/74 (2.7%)	0.211	1.000
	W	5/6 (83.3%)	71/74 (95.9%)	0.272	1.000
248	H	1/6 (16.7%)	0/74 (0.0%)	0.075	1.000
	N	5/6 (83.3%)	73/74 (98.6%)	0.145	1.000
263	D	5/6 (83.3%)	74/74 (100.0%)	0.075	1.000
	E	1/6 (16.7%)	0/74 (0.0%)	0.075	1.000
266	M	1/6 (16.7%)	0/74 (0.0%)	0.075	1.000
	V	5/6 (83.3%)	71/74 (95.9%)	0.272	1.000
288	Q	5/6 (83.3%)	74/74 (100.0%)	0.075	1.000
	R	1/6 (16.7%)	0/74 (0.0%)	0.075	1.000
290	I	5/6 (83.3%)	74/74 (100.0%)	0.075	1.000
	L	1/6 (16.7%)	0/74 (0.0%)	0.075	1.000
332	N	1/6 (16.7%)	20/74 (27.0%)	1.000	1.000
	S	5/6 (83.3%)	42/74 (56.8%)	0.392	1.000
333	K	5/6 (83.3%)	39/74 (52.7%)	0.215	1.000
	N	1/6 (16.7%)	2/74 (2.7%)	0.211	1.000

Transmission patterns

The HOVER-HBV study in DRC identified women who tested HBsAg positive during antenatal screening and performed HBsAg testing on household members (see **Chapter 2**). The study revealed a high HBsAg prevalence (>5%) in households of women tested antenatally for HBsAg[125]. HBsAg positivity was also associated with older age, shared nail clippers, use of street salons and multiple recent sexual partners[125]. Phylogenetic analysis of five households demonstrated that sequences in four households clustered together (Households 1-4, **Figure 6.13**). In Household 2 three distinct strains were identified, 3 individuals had genotype A infection, and one genotype D, and one genotype E, implying transmission from outside the household in the individuals with genotype D and E infection.

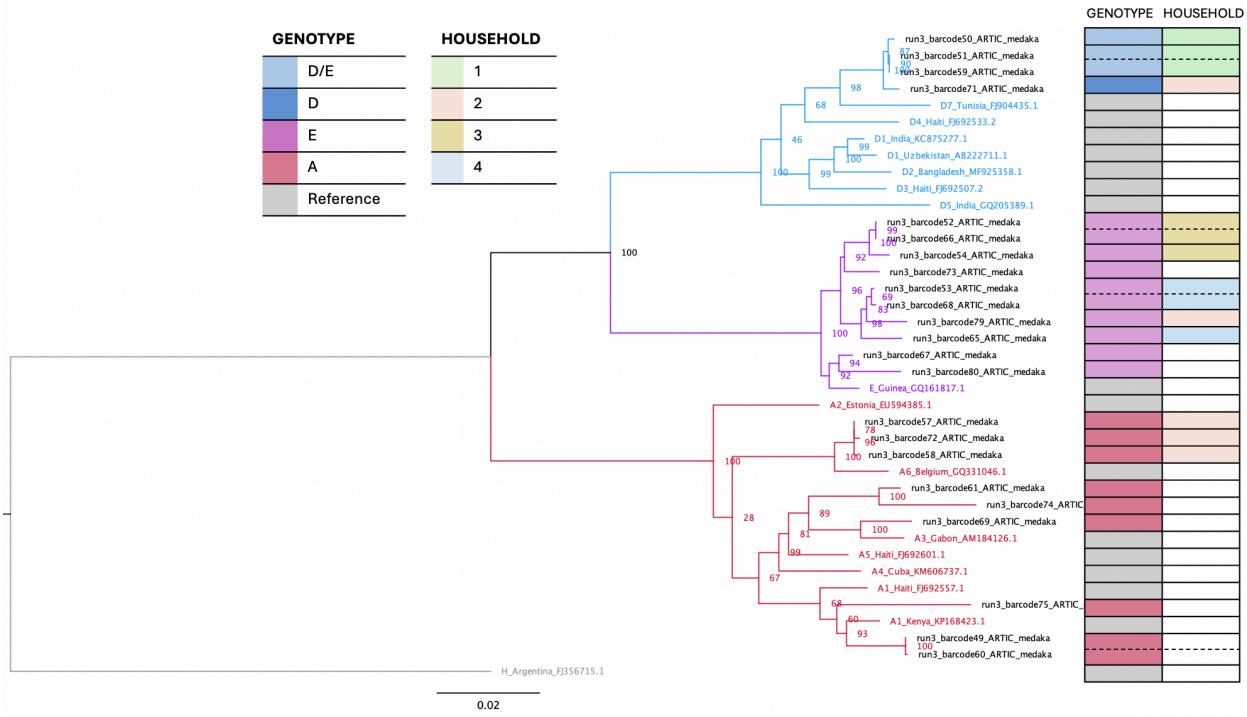


Figure 6.13. Maximum likelihood phylogenetic tree showing genotype and household clustering in the Democratic Republic of Congo with genotype A, D and E reference strains[53]**.** Cells divided by a dotted line indicate sequences from the same individual at different time points (DBS and venous blood draw). Reference strain names are coloured on the tree, with a grey fill in the “Genotype” column. Bootstrap values displayed to the right of nodes. The lengths of the horizontal bars indicate the number of nucleotide substitutions per site.

Clustering of sequences was observed within households in Kenya, indicating that transmission within households occurs in this cohort (**Figure 6.14**). Examples of corresponding family trees produced by Louise Downs are shown below (**Figure 6.15**). The full dataset is required before drawing conclusions about routes of transmission, this will be performed by Louise Downs when the remaining samples have been sequenced.

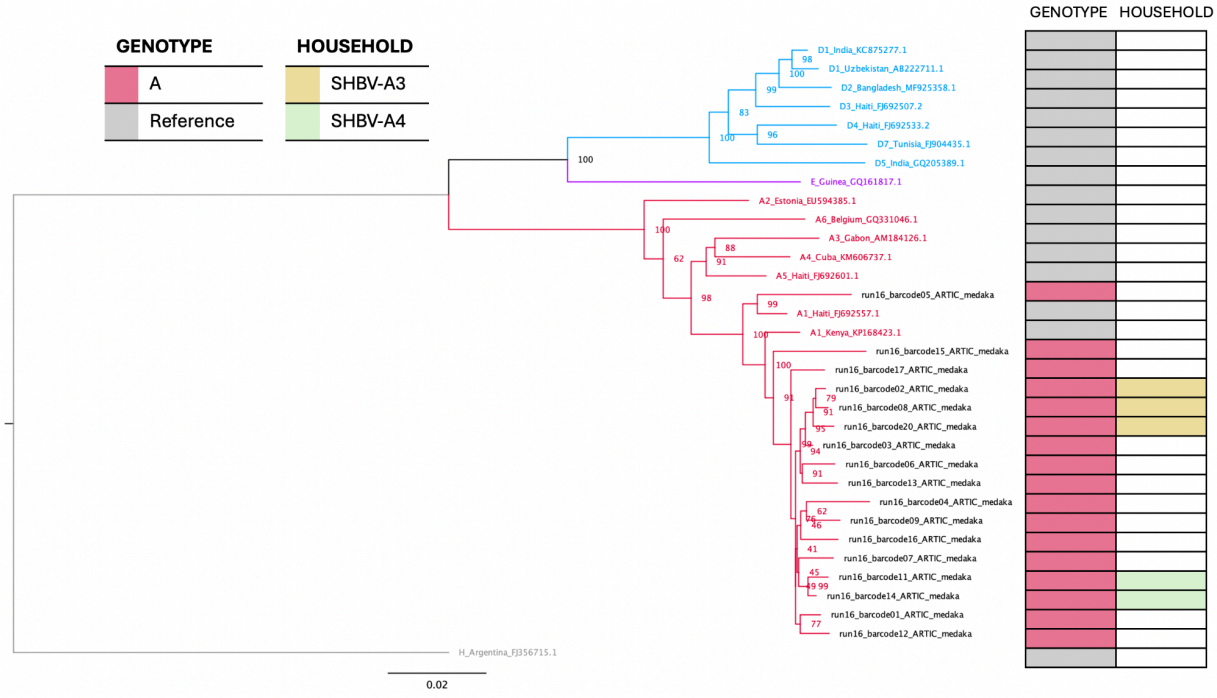


Figure 6.14. Maximum likelihood phylogenetic trees showing genotype and household clustering in Kenya with genotype A, D and E reference strains[53]**.** Reference strain names are coloured on the tree, with a grey fill in the "Genotype" column. Bootstrap values displayed to the right of nodes. The lengths of the horizontal bars indicate the number of nucleotide substitutions per site.

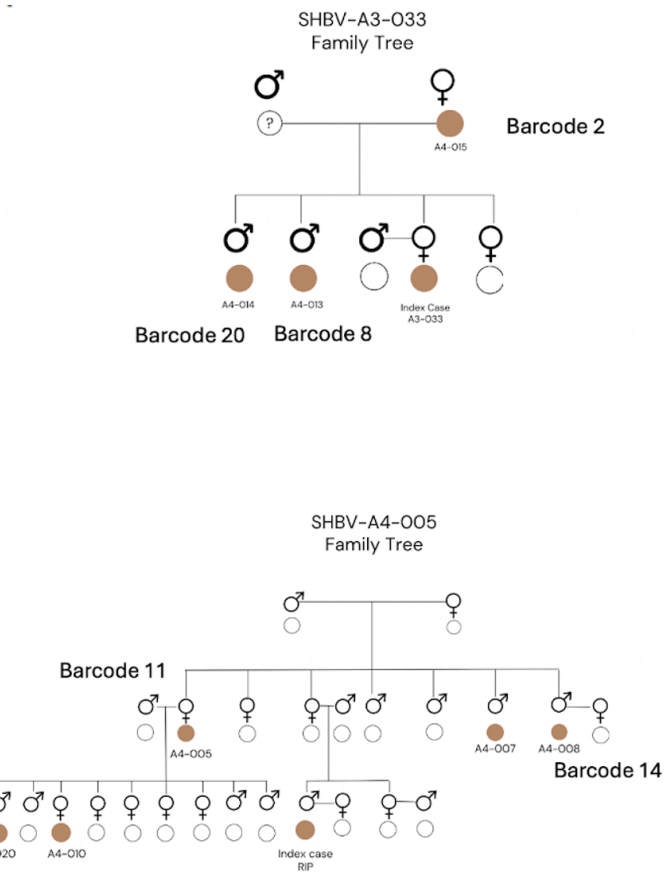


Figure 6.15. Family trees for two households in Kenya highlighted in the phylogenetic tree in Figure 6.14. Circles indicate HBsAg status, individuals who tested HBsAg positive are coloured brown, individuals who tested HBsAg negative are clear. 'Barcode' labels refer to the barcode assigned during WGS of samples from HBsAg positive individuals and correspond to tip labels on the phylogenetic tree in **Figure 6.14** above. The family trees were produced by Louise Downs.

Vaccine escape

One individual with a genotype A HBV mono-infection had the VEM D144E[277,278] (Sample ULIDS-0610, Uganda) and another had the 120T[279] (Sample 218, South Africa), associated with decreased HBsAg expression.

Viral polymorphisms linked to clinical outcomes

Individuals with CHB in Kenya and Uganda had TE performed alongside viral sequencing. I compared the proportion of sequences containing polymorphisms/deletions previously hypothesised to be linked with advanced disease/cirrhosis ((**Table 6.10**)[39,40,280]) in those with/without evidence of fibrosis/cirrhosis (TE threshold 7kPa). The S13T pre-core mutation was enriched in individuals with TE > 7kPa, no other polymorphisms were significantly enriched.

Table 6.10. Frequency of HBV polymorphisms previously associated with cirrhosis identified in consensus sequences derived by Nanopore sequencing among patients with and without a Transient Elastography score > 7kPa (F2 fibrosis).

Position	Frequency (%) of HBV polymorphism in individuals with TE >7kPa (n = 9)	Frequency (%) of HBV polymorphism in individuals with TE < 7kPa (n = 38)	p-value	Adjusted p-value
HbSAg polymorphisms and deletions				
W172* (rtA181T in Pol)	0/9 (0.0%)	0/38 (0.0%)	NA	NA
W182*	0/9 (0.0%)	0/38 (0.0%)	NA	NA
L216*	0/9 (0.0%)	1/35 (2.9%)	1	1
S stop codon	0/9 (0.0%)	0/38 (0.0%)	NA	NA
PreS1 deletion	0/9 (0.0%)	0/33 (0.0%)	NA	NA
PreS2 deletion	1/9 (11.1%)	5/33 (15.2%)	1	1
BCP region, overlapping with X protein changes				
T1753C (I127T in X protein)	1/8 (12.5%)	0/36 (0.0%)	0.182	1
A1762T (K130M in X protein)	3/8 (37.5%)	7/36 (19.4%)	0.355	1
G1764A (V131I in X protein)	5/8 (62.5%)	9/36 (25.0%)	0.087	1
C1766T (F132Y/I/R in X protein)	2/8 (25.0%)	5/36 (13.9%)	0.593	1
T1768A	1/8 (12.5%)	4/36 (11.1%)	1	1
Double mutation (A1762T + G1764A)	3/8 (37.5%)	7/36 (19.4%)	0.355	1
Triple mutation (T1753C + A1762T + G1764A)	1/8 (12.5%)	0/36 (0.0%)	0.182	1
Pre-core region polymorphisms				
S13T	**7/8 (87.5%)**	**3/30 (10.0%)**	**< 0.001**	**0.001**
W28* (also reported as G1896A)	0/8 (0.0%)	0/30 (0.0%)	1	1
G29D (also reported as G1899A)	3/8 (37.5%)	1/30 (3.3%)	0.024	0.427

Stop mutation indicated by *

6.5.4 Conclusion

I have successfully used my HEP-TILE nanopore protocol to generate HBV sequences across four settings, comprising a range of different genotypes. I have contributed significantly to the number of publicly available sequences from DRC and Uganda, and added to the growing set of public sequences from Kenya and South Africa. The full sequencing workflow is now established at KEMRI Wellcome Trust Research Programme (KWTRP) in Kenya.

Using a whole genome approach allowed me to identify genotypes/subgenotypes in addition to recombinants. I confirmed the presence of previously identified D/E recombinants in Uganda, but also identified D/E recombinants in DRC and A/D and A/D/E recombinants in Uganda, which have not been described previously. These may have been missed in previous studies due to the use of partial genome sequencing, or use of techniques such as INNO-LiPA, a line probe based assay which cannot distinguish between mixed and recombinant infections.

I observed transmitted 3TC/ETV RAMs in NA-naive individuals in Uganda (3.1% sequences), but not across the three other sites, this may be due to wider use of 3TC for HIV treatment in the population, or prior undocumented 3TC exposure in the sequenced individuals. There were no RAM combinations associated with TFV resistance in NA-naive individuals across any of the settings. I identified RT polymorphisms in NA-experienced individuals with VBT on TFV, one of which (rt125E) was enriched in NA-experienced individuals compared to NA-naive individuals. The significance of this polymorphism individually and in combination could be explored further using an in vitro phenotypic assay (in development, Khadija Said, Francis Crick Institute). Several putative RAMs e.g. rt126Y, rt153M, rt269L are present in the

wild type reference strain for some subgenotypes; further work is required to determine whether this leads to different barriers to resistance between (sub)genotypes. This has implications for the design of in vitro models of HBVDR, where the genotype backbone should be considered.

Preliminary analysis of data from DRC and Kenya highlights the potential use of WGS in investigating transmission patterns. Analysis of the full sequencing datasets alongside family trees will be performed by Camille Morgan and Louise Downs. Previous studies of household transmission have highlighted the role of both mother to child, father to child and child to child transmission in addition to transmission from sources external to the household [281–285]. Improving understanding of transmission patterns could inform strategic deployment of public health prevention measures, for example testing and vaccination strategies.

The analysis of this data is challenging due to small sample sizes and differences between the clinical metadata collected between settings. The main limitations include a variability in genome coverage from Uganda in particular, as 41% of samples had a VL <2.5 $\log_{10}$ (assay lower limit of quantification 10 IU/ml). The most comprehensive clinical follow up was available for the Uganda setting, my original aim for Uganda data was to perform longitudinal follow up of 220 individuals with serial samples over the 12 year period, however this was not possible due to loss to follow up and sample storage constraints only samples from 2011 and 2023 were available. Numbers of individuals in Uganda with non- genotype A infection followed up longitudinally in 2023 (5/25, with only 1 A/D, 1 A/D/E and 3 D/E) were too small to compare genotype with virological, biochemical or clinical outcomes. The ability of my analysis to detect novel TFV RAMs was limited by the small sample size (n = 6 individuals with VBT on TFV) and by limited clinical information about TFV adherence. HCC status was

not evaluated in these cohorts, therefore I could not evaluate the impact of viral polymorphisms on the development of HCC.

6.6. Discussion

In conclusion, the WHO 2024 HBV treatment guidelines make important progress in facilitating access to treatment for people living with CHB. However, the current lack of accurate, low-cost diagnostics to identify individuals with liver fibrosis and cirrhosis raises concerns these criteria may miss individuals who would benefit from intervention, particularly in settings where TE is not available. By evaluating and comparing treatment stratification methods, this chapter aims to contribute to discussions around efficient management of liver disease and improving health outcomes in resource-limited settings. My work emphasizes the need for longitudinal research cohorts and biobanks with comprehensive clinical metadata in Africa to clarify which individuals benefit most from treatment, which biomarkers are the most appropriate predictors of disease and treatment response, and to evaluate the practicalities and affordability of different stratification approaches.

Future work will link HBV WGS data to metadata on clinical outcomes in a larger multi-centre dataset to investigate the significance of genotype and viral polymorphisms on development of cirrhosis and HCC. As highlighted in **Chapter 5**, longitudinal studies with robust clinical metadata combined with sequencing at baseline and in those who have VBT will help determine risk factors for HBVDR (including the role of minority variants) and poor clinical outcomes e.g. cirrhosis and HCC. Although WGS has the benefit of being able to evaluate the whole genome in a single assay, HEP-TILE does not allow haplotypes extending across amplicons to be called. Long read sequencing (Nanopore or PacBio) of a single ~1kb amplicon covering the RT gene could be used to determine viral resistance haplotypes within

the RT gene[286] in addition to single genome amplification. Similarly, strategic use of longer amplicons and long read sequencing, will assist identification of recombinant strains which is challenging with short read sequencing, although due to known limitations in sensitivity this is likely to only be of use in samples with VL > $4\log_{10}$ IU/ml. Future work will build on existing collaborations, with capacity building in the WHO African region and bi-directional knowledge transfer.

While the WHO 2024 treatment criteria offer an opportunity to expand access to care, they also highlight the urgent need for contextually appropriate diagnostic strategies tailored to available resources. Addressing these challenges is vital for improving outcomes in the management of hepatitis B and achieving global health equity.

Chapter 7: Discussion

7.1 Introduction

My thesis aims to impact the clinical and public health management of SARS-CoV-2 and HBV by developing and implementing WGS as a translational tool. For SARS-CoV I investigate the utility of WGS for investigating transmission within healthcare settings. For HBV, I focus on expanded treatment guidelines and the potential risk of HBVDR using WGS across four settings across Africa to improve our understanding of genetic epidemiology, HBVDR and transmission.

7.2 Summary of work, impact and limitations

7.2.1 SARS-CoV-2: WGS insights into nosocomial transmission

My work in **Chapter 3** highlights the role of genomic epidemiology in characterising and managing nosocomial SARS-CoV-2 transmission. I highlight the limitations of traditional epidemiological classifications of nosocomial outbreaks, especially during periods of high SARS-CoV-2 incidence. I demonstrate how WGS can be used to provide a more accurate picture of transmission within the hospital setting, revealing that only a small proportion (7%) of individual cases lead to onward transmission, and that superspreading events drive most (80%) of nosocomial cases. Notably, epidemiological data appeared to overestimate the role of healthcare staff in spreading infections. This more nuanced understanding prompted a renewed focus on admission screening and rapid action in response to new inpatient cases, including whole ward screening, particularly in mixed hospital sites. On a public health level, I highlight critical issues with the UK national reporting criteria for nosocomial infections, revealing that nosocomial incidence was underestimated nationally based on the definitions

in use at the time and highlighting the need for new definitions to better capture the true burden of hospital-acquired SARS-CoV-2 (definitions were not changed).

In this study I established a fast, high-throughput SARS-CoV-2 sequencing pipeline at the John Radcliffe Hospital, Oxford, this capability allowed OUH to contribute actively to the UK's public health response. Sequencing results were integrated into the national system for SARS-CoV-2 surveillance, as well as being added to the Global Pathogen Analysis Service (GPAS), to aid the surveillance and tracking of SARS-CoV-2 globally. My data allowed immediate responses to requests from UKHSA about whether potential outbreaks involved VOCs; in two cases I was able to confirm that single individuals with VOCs were linked to known community outbreaks, but were not linked to current hospital outbreaks. The in-house sequencing capability led to a number of other research collaborations, facilitating rapid pace translational research assessing the impact of viral mutations on escape from natural and vaccine induced immunity (collaboration with Gavin Screaton)[287,288]. I also used genomic data to confirm/exclude reinfection in individuals with chronic kidney disease and in a study of antibody responses in those with chronic myeloid neoplasms[289,290].

I demonstrate the utility of batched retrospective on-site WGS. As demonstrated by the COG-HOCI study, the utility of near real time sequencing (<5 days) in a pandemic setting is constrained by the case load and resource limitations in both healthcare and research contexts and proximity of diagnostic and sequencing labs[207]. Real time WGS would need to be integrated within a responsive system with short turnaround times, sufficient clinical bandwidth, and adequate hospital resources for cohorting or isolation to allow clinical benefit to be realised. The main limitation of this study was the challenge of missing data. Nosocomial transmission could not always be ruled out due to incomplete data from undiagnosed or unsequenced individuals.

7.2.2 Improved techniques for HBV WGS

I developed three distinct viral enrichment methods for WGS of HBV: HEP-TILE, probe-based capture, and gPCR-RCA (**Chapter 4**). The HEP-TILE method offers a significant improvement in sensitivity compared to previously used sequencing methods. A key strength of my analysis was the inclusion of mock samples, which enabled assessment of performance across rare HBV genotypes (F-J) that are less prevalent in our cohorts but present globally. This ensures global applicability of the protocol, thereby enhancing the utility across diverse HBV populations.

By developing and publishing an easy to use, open-access wet lab and bioinformatic pipeline, based on the widely used SARS-CoV-2 ARTIC sequencing protocol, the HEP-TILE protocol will be straightforward to adopt worldwide due to the global expansion of expertise and infrastructure in amplicon-based Nanopore sequencing during the pandemic, I have already transferred the methods to KEMRI-Wellcome Trust Research Programme (Kenya) and Uganda Virology Research Institute (Uganda) with future plans to take to South Africa (OXSAHEP cohort) and Democratic Republic of Congo (collaboration with Professor Peyton Thompson). The UKHSA blood borne virus unit has also adopted my protocol for HBV WGS. These methods will support the generation of more globally representative HBV sequence data, which is crucial for accurately mapping global HBV molecular epidemiology, shedding light on persistence and pathogenesis, and understanding treatment responses and clinical outcomes more comprehensively. The flexibility in sequencing from a venous blood draw as well as DBS removes traditional barriers associated with sample storage and transport, making the protocols more accessible to resource-limited settings. A limitation of the sequencing method development was the restricted availability of HBV samples and limited sample volumes, which constrained the number of comparisons between methods.

7.2.3 HBV drug resistance knowledge gaps and systematic review

In **chapter 5** I identify and address clinical and public health questions surrounding HBVDR in the context of expanded WHO treatment guidelines. I worked with the WHO to write a narrative review in which I outline the current HBVDR landscape, discuss HBVDR as a potential threat to global HBV elimination efforts, identify knowledge gaps and suggest a roadmap highlighting seven areas for action (manuscript accepted by Lancet Microbe). This framework aims to advocate for and therefore improve global surveillance of HBVDR. I fill one of the evidence gaps by performing the first systematic review and meta-analysis of ETV and TFV resistance rates. I show that TFV exhibits a high genetic barrier to resistance compared to ETV, which is more prone to resistance, especially in treatment-experienced patients. These findings have implications for clinical decision-making, guiding clinicians and guideline committees on choice of first-line therapy and highlighting the importance of treatment history, and were presented to a WHO-hosted webinar for World Antimicrobial Awareness Week (November 2024).

The meta-analysis has limitations, primarily due to the small number of trials in each category, heterogeneity in clinical methods and patient characteristics and representation of certain global regions and populations. Many studies were clinical trials excluding key high-risk populations, such as individuals with inconsistent adherence, co-infections (HIV, HCV, HDV), or comorbidities, and populations with high mobility (e.g. migrants, travellers, those experiencing homelessness), as well as those who are pregnant or breastfeeding. Additionally, the global data coverage was uneven, with little data from sub-Saharan Africa, where barriers to treatment access and HIV/HBV co-endemicity might increase resistance risk. This limitation could lead to an underestimation of global HBVDR prevalence,

particularly in the WHO African region and in key populations. I emphasise the inadequate representation of populations most at risk for HBVDR and highlight the challenge of working with incomplete metadata and inconsistent definitions and algorithms for identifying HBVDR. This informs the focus areas for future research, particularly in improving data collection for risk populations and refining the interpretation of resistance data within various global contexts.

7.2.4 Application of new WHO HBV treatment guidelines in Uganda

In **Chapter 6**, I apply the new WHO treatment guidelines to a cohort in rural Uganda, offering real-world insights into the challenges and limitations faced when applying these guidelines in resource-limited settings. I demonstrate how diagnostic availability influences our ability to identify patients who might benefit from HBV treatment, due to the low sensitivity of APRI for predicting liver fibrosis/cirrhosis in this cohort. I question whether alternate scoring systems or adjusted thresholds developed on data from African cohorts might yield better predictive accuracy in such settings. The results help guide resource allocation in Uganda and provide the WHO with critical feedback on the limitations of its treatment guidelines when applied in the settings for which they have been designed. Finally I apply the HBV WGS methods developed in **Chapter 4**, to identify genotypes and establish HBVDR prevalence across four settings in Africa.

My work on treatment eligibility demonstrates the need for longitudinal larger scale studies in Africa, providing a more representative understanding of treatment efficacy and HBV outcomes across diverse populations. Additional research should examine the predictors of cirrhosis and HCC in African cohorts, exploring the feasibility, cost-effectiveness and acceptability of broader "treat-all" strategies which could be used if access to diagnostics is limited. The primary limitation is the small cohort and infrequent follow-up, with assessments

occurring only twice over ten years, as opposed to the recommended biannual monitoring. This could lead to an underestimation of treatment eligibility based on serial HBV DNA or ALT measures, given the potential for missed fluctuations in HBV DNA levels. However, this limitation itself provides insights into the logistical and practical challenges in regions where regular follow-up is neither feasible nor affordable. The high rate of loss to follow-up further reflects the difficulties of sustained cohort studies in mobile populations and underscores the need for enhanced patient retention and biobank resources in future studies.

7.2.5 HBV WGS in the WHO African region

My work on expanding WGS to different settings across Africa is part of an ongoing multi-centre study to investigate the impact of host and viral genetics on clinical outcomes. I highlight the utility of HBV WGS in determining genotype, subgenotype, recombinants, HBVDR and understanding transmission patterns, identifying novel recombinant strains in Uganda and DRC. The main limitations of the work are the small sample size, which when combined with the restricted range of genotypes, low percentage of individuals on treatment, and low prevalence of HBVDR limited further analysis on correlates of treatment success. Only 5 of the original 220 person cohort in Uganda had serial samples available for sequencing, preventing the evaluation of viral evolution and relationship between viral genome and clinical outcomes that was initially intended, nevertheless the 42 whole genome sequences produced here represent an important contribution to the literature as only one HBV whole genome sequence from Uganda was available on GenBank previously. Future work will link HBV WGS data to metadata on clinical outcomes in a larger dataset to allow risk factors for HBVDR to be identified and significance of genotype and viral polymorphisms on clinical outcomes to be determined.

7.3 Challenges with SARS-CoV-2 and HBV research

A shared theme in my SARS-CoV-2 and HBV research is the challenge of achieving adequate data representation. This is firstly due to laboratory challenges of difficult-to-sequence low VL samples, the presence of host background material and undiagnosed individuals within a population. Ultimately my experience of SARS-CoV-2 sequencing informed my approaches to sequencing HBV low VL samples, leading to a collaboration with the University of Birmingham team who designed the ARTIC SARS-CoV-2 primers, using a tiled amplicon approach with short amplicons for both pathogens. Viral diversity presents a challenge for tiled amplicon sequencing; primer schemes need to be updated as the sequenced diversity of a pathogen increases. The ARTIC collaboration released v4 SARS-CoV-2 primers as VOCs emerged[291]), similarly as more HBV WGS are generated, additional primers may need to be added to the HEP-TILE primer pool to account for newly sequenced diversity. Secondly there is the challenge of representing the communities and populations at risk. In the context of SARS-CoV-2 staff testing, certain staff groups found it hard to access asymptomatic and symptomatic staff testing, due to shift work, concerns about loss of pay/employment, language barriers, and digital exclusion, all compounded by high staff turnover. In the context of HBV there is a challenge of global representation of clinical and sequencing data, and a challenge of generating data representing vulnerable HBV populations who may be harder to reach and follow up longitudinally.

The COVID-19 pandemic illustrated the impact of focused, high-intensity global research efforts on the rapid development and licensing of vaccines and therapeutics. The speed, scale and prioritisation of SARS-CoV-2 research contrasted markedly with the HBV research landscape. During the pandemic, the rapid convergence of institutions, researchers, funders,

the public and publishers on a single pathogen led to a fast-paced, intense period in which collaboration and scientific output surged. In contrast, HBV has historically been neglected in research funding and global health agendas[33]. The COVID-19 pandemic further slowed HBV research due to the suspension of many clinical and research programmes and has impacted on hepatitis B and C elimination[292]. Resuming HBV research post-pandemic has required time to rebuild momentum and re-establish collaborations.

7.4 Future recommendations and next steps

To address the need for better global representation in HBV sequencing data, I am working with members of the Matthews group and collaborating with other international teams in Uganda, Kenya, South Africa, Malawi and the Democratic Republic of Congo to establish HBV WGS. By improving global representation in sequencing real-world datasets linked with high-quality clinical metadata including diverse populations with risk factors for HBVDR, our joint objectives are:

1) to investigate the effects of viral genetic variation on clinical outcomes such as treatment response, liver cirrhosis and HCC with the potential for personalized care including risk stratification and drug-resistance monitoring;
2) to investigate the genetic basis for treatment non-response and HBVDR and its practical applications for patient care;
3) to produce whole genome sequences to use in *in vitro* models for studying tenofovir resistance;
4) to investigate how host-virus interactions influence disease outcomes;
5) to use WGS to investigate mother-to-child transmission and household transmission patterns and drug resistance in Kenya and DRC.

The WHO is in the process of writing an integrated Global Action Plan (iGAP) to address drug resistance across HIV, viral hepatitis, and sexually transmitted infections. This iGAP will expand on the existing five-year HIV drug resistance GAP, developed by WHO's HIVResNet. The iGAP will incorporate HBVDR as a key focus area, aiming to identify resistance risk factors, monitor HBVDR prevalence and geographic distribution, and evaluate the clinical impact of HBVDR, drawing on previous experience with HIV. Development of an HBVDR database as a global repository of HBVDR data, with de-identified individual-level epidemiological information linked to HBV genome sequences, would allow prevalence of known and novel RAMs leading to acquired and transmitted HBVDR to be mapped and tracked.

Emerging HBV therapies aim to achieve a functional cure (sustained HBsAg loss +/- seroconversion to anti-HBs and undetectable HBeAg and serum HBV DNA after a finite course of therapy) by targeting different steps in the HBV lifecycle or acting as immune modulators. HBV WGS will be essential to monitor for emergence of escape from HB-specific B or T cell mediated therapies, and direct acting antiviral strategies such as capsid assembly modulators and small interfering RNAs.

Ultimately, progression towards the 2030 elimination targets requires sustained investment in the prevention, diagnosis and treatment of CHB. Low diagnosis and treatment coverage remain major barriers to reducing the morbidity and mortality associated with CHB globally; there is a critical need to re-think diagnostic strategies, models of care and improve linkage to care, in order to tackle HBV as a global health threat.

Appendices

Appendix 1: Publications and presentations

In April 2020, I was involved in establishing a framework for a longitudinal cohort of Oxford University Hospitals (OUH) staff, to track infection and immunity, generating SARS-CoV-sequences and antibody data to advance our understanding of the clinical and molecular epidemiology of SARS-CoV-2 and inform surveillance, diagnostics, and IPC. This resulted in a number of publications in addition to the nosocomial transmission study I present in **Chapter 3**[166].

My first study examined differential occupational risks among HCWs, showing higher COVID-19 rates in COVID-19-facing areas, with acute medicine staff facing greater risks, compared to intensive care unit staff who were better protected by PPE bundles. Black, Asian, and minority ethnic (BAME) staff, as well as porters and cleaners, were disproportionately affected. This research led to the implementation of 10 key IPC measures, including enhanced PPE use, social distancing, and risk assessments for BAME staff.

I coordinated the Respiratory Diagnostic Kit Evaluation (‘Red Kite’) study (Research Ethics Committee reference: 19/NW/0730; North West-Greater Manchester South Research Ethics Committee). This fed into the hospital arm of the UK COVID-19 Lateral Flow study, a head-to-head-benchmark comparison of SARS-CoV-2 lateral flow tests[293]; I contributed the retrospective secondary care evaluation (Phase 3a) and community field service evaluations (Phase 4) of a range of lateral flow tests.

Next, I explored antibody dynamics and immunity following infection and vaccination. One study[294] tracked antibody responses in 3,276 HCWs, showing a decline in anti-nucleocapsid antibodies over time, but stable anti-spike antibodies. Another paper published in *New England Journal of Medicine*[295] demonstrated that prior infection with SARS-CoV-2 conferred a 6-month protection against reinfection, evidenced by the reduced incidence of symptomatic infections among seropositive HCWs; this was the first paper in the world to demonstrate a degree of immunity after natural infection.

Following the rollout of vaccines in December 2020, I found that vaccination generated robust anti-spike antibody responses, with prior infection leading to significantly higher antibody levels post-vaccination[296]. A subsequent study[297] confirmed that two doses of either Pfizer-BioNTech or Oxford-AstraZeneca vaccines provided similar levels of protection as natural immunity, reducing symptomatic and PCR-positive infections and suggesting that one dose may suffice in individuals with prior infection.

I established an in-house SARS-CoV-2 sequencing capability which led to a number of other research collaborations, facilitating rapid pace translational research assessing the impact of viral mutations on escape from natural and vaccine induced immunity[287,288,298] and studying immune responses and reinfection in immunocompromised populations[289,290].

I have published my HBV work in this thesis in two manuscripts[100,234] and two pre-prints[133,208], I have two papers accepted awaiting publication at Lancet Microbe (WHO HBVDR review) and Scientific Reports (HBV WGS methods), and have presented the work at a number of international conferences[209–211,259].

Appendix 2: Sample metadata

Supplementary Table 1: HBV samples used for rolling circle amplification. I used these samples in Chapter 4, section 4.4: HBV rolling circle amplification. Samples given abbreviated names (S1-8) to facilitate figure labelling.

Participant ID	Figure label	VL	Cohort	Host depletion	Extraction	Input vol	Elution vol
Host depletion comparison (chapter 4, section 4.2.4) and phi29/equiphi comparison (chapter 4, 4.2.5)							
HEP1409 (Med 560)	S1	7.47	UK	No v mnase v neb	KF Magmax	360	30
CARTH03537	S2	5.69	UK	No v mnase v neb	KF Magmax	360	30
HEP1522	S3	3.40	UK	No v mnase v neb	KF Magmax	360	30
HEP1745	S4	1.65	UK	No v mnase v neb	KF Magmax	360	30
Primer number							
HEP1152	S5	3.90	UK	mnase	KF Magmax	360	30
HEP1361 (Med 1331)	S6	5.00	UK	mnase	KF Magmax	360	30
Generation of circular template							
HEP1361 (Med 1331)	S7	8.23	UK	mnase	KF Magmax	360	30
HEP1249 (Med 1318)	S8	5.50	UK	mnase	KF Magmax	360	30
JG83	S9	7.12	UCL	mnase	KF Magmax	360 x2	35 x2
1103	S10	5.69	SA	mnase	KF Magmax	360 x3	35 x3

Supplementary Table 2: HBV samples used for takara library prep and capture. I used these samples in Chapter 4, section 4.5.2.

Participant ID	VL	Cohort	Host depletion	Extraction	Input vol	Elution vol
HEP1711	4.07	UK	none	Nuclisens EasyMag	1000	50
HEP1546	3.89	UK	none	Nuclisens EasyMag	1000	50
HEP1452	3.52	UK	none	Nuclisens EasyMag	1000	50
HEP1522	3.36	UK	none	Nuclisens EasyMag	1000	50
HEP1489	3.19	UK	none	Nuclisens EasyMag	1000	50
HEP1191	3.18	UK	none	Nuclisens EasyMag	1000	50
HEP1385	2.89	UK	none	Nuclisens EasyMag	1000	50
HEP1028	2.88	UK	none	Nuclisens EasyMag	1000	50
HEP1018	2.78	UK	none	Nuclisens EasyMag	1000	50
HEP1462	2.77	UK	none	Nuclisens EasyMag	1000	50
HEP1165	2.76	UK	none	Nuclisens EasyMag	1000	50
HEP1401	2.76	UK	none	Nuclisens EasyMag	1000	50
HEP1622	2.68	UK	none	Nuclisens EasyMag	1000	50
HEP1328	2.64	UK	none	Nuclisens EasyMag	1000	50
HEP1190	2.55	UK	none	Nuclisens EasyMag	1000	50
HEP1341	2.52	UK	none	Nuclisens EasyMag	1000	50
HEP1433	2.49	UK	none	Nuclisens EasyMag	1000	50
HEP1167	2.45	UK	none	Nuclisens EasyMag	1000	50
HEP1099	2.02	UK	none	Nuclisens EasyMag	1000	50
HEP1149	1.98	UK	none	Nuclisens EasyMag	1000	50

Supplementary Table 3: HBV samples used for NEB library prep and capture. I used these samples in Chapter 4, section 4.5.2.

Participant ID	VL	Cohort	Host dep	Extraction	Input vol	Elution vol
1113	> 7.2	SA	mnase	KF Magmax	250 x2	50 x2
1162	7.00	SA	mnase	KF Magmax	250 x2	50 x2
1143	6.73	SA	mnase	KF Magmax	250 x2	50 x2
1191	6.62	SA	mnase	KF Magmax	250 x2	50 x2
1114	6.31	SA	mnase	KF Magmax	250 x2	50 x2
251	6.29	SA	mnase	KF Magmax	250 x2	50 x2
1125	6.17	SA	mnase	KF Magmax	250 x2	50 x2
200	6.09	SA	mnase	KF Magmax	250 x2	50 x2
1175	6.06	SA	mnase	KF Magmax	250 x2	50 x2
1103	5.69	SA	mnase	KF Magmax	250 x2	50 x2
218	5.59	SA	mnase	KF Magmax	250 x2	50 x2
1119	5.53	SA	mnase	KF Magmax	250 x2	50 x2
1131	5.47	SA	mnase	KF Magmax	250 x2	50 x2
1116	5.41	SA	mnase	KF Magmax	250 x2	50 x2
1105	5.36	SA	mnase	KF Magmax	250 x2	50 x2
1115	5.34	SA	mnase	KF Magmax	250 x2	50 x2
1132	5.15	SA	mnase	KF Magmax	250 x2	50 x2
1147	5.09	SA	mnase	KF Magmax	250 x2	50 x2
211	5.01	SA	mnase	KF Magmax	250 x2	50 x2
1108	4.64	SA	mnase	KF Magmax	250 x2	50 x2
1174	4.44	SA	mnase	KF Magmax	250 x2	50 x2
1134	4.32	SA	mnase	KF Magmax	250 x2	50 x2
1170	4.21	SA	mnase	KF Magmax	250 x2	50 x2

Supplementary Table 4: HBV samples used for HEP-TILE proof of principle work. I used these samples in Chapter 4, section 4.6 and 4.7.

Participant ID	VL (log 10 IU/ml)	Cohort	Host dep	Extraction	Input vol	Elution vol
HEP-TILE nanopore, illumina and capture						
1113	> 7.20	SA	mnase	KF Magmax	250 x2	50 x2
1143	6.73	SA	mnase	KF Magmax	250 x2	50 x2
1191	6.62	SA	mnase	KF Magmax	250 x2	50 x2
251	6.29	SA	mnase	KF Magmax	250 x2	50 x2
200	6.09	SA	mnase	KF Magmax	250 x2	50 x2
1175	6.06	SA	mnase	KF Magmax	250 x2	50 x2
218	5.59	SA	mnase	KF Magmax	250 x2	50 x2
1119	5.53	SA	mnase	KF Magmax	250 x2	50 x2
1116	5.41	SA	mnase	KF Magmax	250 x2	50 x2
1115	5.34	SA	mnase	KF Magmax	250 x2	50 x2
HEP-TILE nanopore only						
900219N	8.23	Uganda	none	Qiagen MinElute	200	45
906667N	8.23	Uganda	none	Qiagen MinElute	200	45
907906N	7.60	Uganda	none	Qiagen MinElute	200	45
900603N	7.50	Uganda	none	Qiagen MinElute	200	45
906116N	7.02	Uganda	none	Qiagen MinElute	200	45
ULIDS-0973	6.13	Uganda	none	Qiagen MinElute	200	45
ULIDS-0610	4.64	Uganda	none	Qiagen MinElute	200	45
907395N	4.05	Uganda	none	Qiagen MinElute	200	45
901116N	3.90	Uganda	none	Qiagen MinElute	200	45
ULIDS-0434	3.72	Uganda	none	Qiagen MinElute	200	45
ULIDS-0001	3.63	Uganda	none	Qiagen MinElute	200	45
903818N	3.61	Uganda	none	Qiagen MinElute	50	45
ULIDS-0457	3.53	Uganda	none	Qiagen MinElute	200	45

908210N	3.36	Uganda	none	Qiagen MinElute	200	45
ULIDS-0461	3.14	Uganda	none	Qiagen MinElute	200	45
907856N	3.12	Uganda	none	Qiagen MinElute	200	45
ULIDS-0943	2.97	Uganda	none	Qiagen MinElute	400	45
ULIDS-0595	2.93	Uganda	none	Qiagen MinElute	400	45
ULIDS-0955	2.88	Uganda	none	Qiagen MinElute	400	45
905960N	2.29	Uganda	none	Qiagen MinElute	400	45
HEP1363	5.64	UK	none	Qiagen MinElute	200	45
HEP1745	5.50	UK	none	Qiagen MinElute	200	45
HEP1740	4.82	UK	none	Qiagen MinElute	200	45
HEP0709	4.51	UK	none	Qiagen MinElute	200	45
HEP1749	4.49	UK	none	Qiagen MinElute	200	45
HEP1740	4.40	UK	none	Qiagen MinElute	200	45
HEP1727	4.10	UK	none	Qiagen MinElute	200	45
HEP1763	3.79	UK	none	Qiagen MinElute	200	45
HEP1745	3.73	UK	none	Qiagen MinElute	200	45
HEP1747	3.35	UK	none	Qiagen MinElute	200	45
HEP1571	3.16	UK	none	Qiagen MinElute	200	45
HEP1403	3.01	UK	none	Qiagen MinElute	200	45
HEP1104	2.63	UK	none	Qiagen MinElute	400	45
HEP1752	2.55	UK	none	Qiagen MinElute	400	45
HEP1746	2.52	UK	none	Qiagen MinElute	400	45
HEP1766	2.18	UK	none	Qiagen MinElute	400	45
HEP1681	1.90	UK	none	Qiagen MinElute	400	45
HEP1529	1.62	UK	none	Qiagen MinElute	400	45
HEP1710	1.50	UK	none	Qiagen MinElute	400	45

Supplementary Table 5: HBV samples used for ultra low VL work. I used these samples in Chapter 4, section 4.8.

Participant ID	VL (log 10 IU/ml)	Cohort	Host dep	Extraction	Input vol	Elution vol
Bakeoff_1	2.61	NHSBT	none	Roche High Pure	5000	50
Bakeoff_2	3.60	NHSBT	none	Roche High Pure	5000	50
Bakeoff_3	0.00 (1 IU/ml)	NHSBT	none	Roche High Pure	5000	50
Bakeoff_4	1.46	NHSBT	none	Roche High Pure	5000	50
Bakeoff_5	1.91	NHSBT	none	Roche High Pure	5000	50
Bakeoff_6	1.78	NHSBT	none	Roche High Pure	5000	50
Bakeoff_7	2.90	NHSBT	none	Roche High Pure	5000	50
Bakeoff_8	3.67	NHSBT	none	Roche High Pure	5000	50
Bakeoff_9	2.72	NHSBT	none	Roche High Pure	5000	50
Bakeoff_10	0.00 (neg)	NHSBT	none	Roche High Pure	5000	50
Bakeoff_11	0.00 (1 IU/ml)	NHSBT	none	Roche High Pure	5000	50
Bakeoff_12	1.68	NHSBT	none	Roche High Pure	5000	50
Bakeoff_13	2.23	NHSBT	none	Roche High Pure	5000	50
Bakeoff_14	3.23	NHSBT	none	Roche High Pure	5000	50
Bakeoff_15	3.72	NHSBT	none	Roche High Pure	5000	50
Bakeoff_16	2.53	NHSBT	none	Roche High Pure	5000	50
Bakeoff_17	2.66	NHSBT	none	Roche High Pure	5000	50
Bakeoff_18	1.66	NHSBT	none	Roche High Pure	5000	50
Bakeoff_19	3.00	NHSBT	none	Roche High Pure	5000	50
Bakeoff_20	1.41	NHSBT	none	Roche High Pure	5000	50
Bakeoff_21	1.43	NHSBT	none	Roche High Pure	5000	50
Bakeoff_22	2.10	NHSBT	none	Roche High Pure	5000	50
Bakeoff_23	2.85	NHSBT	none	Roche High Pure	5000	50
Bakeoff_24	0.90	NHSBT	none	Roche High Pure	5000	50

HBV samples from four settings in the WHO African region

I used the DRC DBS samples in Chapter 4, section 4.9. I report the results from all samples in Chapter 6, section 6.5.

Supplementary Table 6: HBV samples from Democratic Republic of the Congo

Participant ID	VL	Cohort	Host dep	Extraction	Input vol	Elution vol
Dried Blood spot (DBS)						
HRB-1028-01-D	5.15	DRC	none	chelex tween	6mm punch	150
HRB-1083-11-D	5.07	DRC	none	chelex tween	6mm punch	150
AVK-2039-enr-D	5.07	DRC	none	chelex tween	6mm punch	150
HRK-2083-02-D	4.39	DRC	none	chelex tween	6mm punch	150
AVK-2016-enr-D	4.36	DRC	none	chelex tween	6mm punch	150
HRB-1082-01-D	3.98	DRC	none	chelex tween	6mm punch	150
AVK-2039-SA-D	3.78	DRC	none	chelex tween	6mm punch	150
HRK-2043-04-D	3.53	DRC	none	chelex tween	6mm punch	150
Serum						
HRB-1013-05-S	Not detected	DRC	none	Qiagen DNA blood midi	2000	110
HRK-2043-04-S	Not detected	DRC	none	Qiagen DNA blood midi	2000	110
HRK-2043-06-S	Not detected	DRC	none	Qiagen DNA blood midi	2000	110
HRB-1028-05-S	8.68	DRC	none	Qiagen DNA blood midi	2000	110
HRB-1083-11-S	8.46	DRC	none	Qiagen DNA blood midi	2000	110
HRB-1028-01-S	8.44	DRC	none	Qiagen DNA blood midi	2000	110
HRK-2083-03-S	7.87	DRC	none	Qiagen DNA blood midi	2000	110
HRB-1082-01-S	7.86	DRC	none	Qiagen DNA blood midi	2000	110
HRK-2083-02-S	7.79	DRC	none	Qiagen DNA blood midi	2000	110
HRK-2043-01-S	6.08	DRC	none	Qiagen DNA blood midi	2000	110
HRK-2043-10-S	5.68	DRC	none	Qiagen DNA blood midi	2000	110

HRB-1082-06-S	4.72	DRC	none	Qiagen DNA blood midi	2000	110
HRB-1006-04-S	4.14	DRC	none	Qiagen DNA blood midi	2000	110
HRK-2043-05-S	3.56	DRC	none	Qiagen DNA blood midi	2000	110
HRK-2081-01-S	3.36	DRC	none	Qiagen DNA blood midi	2000	110
HRB-1034-01-S	3.27	DRC	none	Qiagen DNA blood midi	2000	110
HRK-2027-01-S	3.25	DRC	none	Qiagen DNA blood midi	2000	110
HRB-1083-08-S	3.15	DRC	none	Qiagen DNA blood midi	2000	110
HRK-2016-01-S	2.76	DRC	none	Qiagen DNA blood midi	2000	110
HRK-2043-03-S	2.52	DRC	none	Qiagen DNA blood midi	2000	110
HRK-2043-11-S	2.43	DRC	none	Qiagen DNA blood midi	2000	110
HRK-2089-01-S	2.2	DRC	none	Qiagen DNA blood midi	2000	110
HRK-2057-01-S	2.07	DRC	none	Qiagen DNA blood midi	2000	110
HRK-2083-01-S	1.72	DRC	none	Qiagen DNA blood midi	2000	110
HRK-2089-05-S	1.15	DRC	none	Qiagen DNA blood midi	2000	110

Supplementary Table 7: HBV samples from Uganda

Participant ID	VL	Cohort	Host dep	Extraction	Input vol	Elution vol
900219N	8.23	Uganda	none	Qiagen MinElute	200	45
906667N	8.23	Uganda	none	Qiagen MinElute	200	45
904710N	8.23	Uganda	none	Qiagen MinElute	50	45
901993N	8	Uganda	none	Qiagen MinElute	100	45
905350N	7.94	Uganda	none	Qiagen MinElute	50	45
907906N	7.6	Uganda	none	Qiagen MinElute	200	45
900603N	7.5	Uganda	none	Qiagen MinElute	200	45
906116N	7.02	Uganda	none	Qiagen MinElute	200	45
ULIDS-0973	6.13	Uganda	none	Qiagen MinElute	200	45
ULIDS-0276	5.27	Uganda	none	Qiagen MinElute	200	45
ULIDS-0531	4.95	Uganda	none	Qiagen MinElute	200	45
ULIDS-0610	4.64	Uganda	none	Qiagen MinElute	200	45
907327N	4.55	Uganda	none	Qiagen MinElute	200	45
ULIDS-0092	4.36	Uganda	none	Qiagen MinElute	200	45
907395N	4.05	Uganda	none	Qiagen MinElute	200	45
907273N	3.97	Uganda	none	Qiagen MinElute	200	45
901116N	3.9	Uganda	none	Qiagen MinElute	200	45
903581N	3.8	Uganda	none	Qiagen MinElute	200	45
ULIDS-0025	3.76	Uganda	none	Qiagen MinElute	200	45
ULIDS-0434	3.72	Uganda	none	Qiagen MinElute	200	45
ULIDS-0001	3.63	Uganda	none	Qiagen MinElute	200	45
903818N	3.61	Uganda	none	Qiagen MinElute	50	45
ULIDS-0457	3.53	Uganda	none	Qiagen MinElute	200	45
908210N	3.36	Uganda	none	Qiagen MinElute	200	45
903887N	3.31	Uganda	none	Qiagen MinElute	50	45
904745N	3.29	Uganda	none	Qiagen MinElute	50	45
ULIDS-0461	3.14	Uganda	none	Qiagen MinElute	200	45

907856N	3.12	Uganda	none	Qiagen MinElute	200	45
ULIDS-0428	3.09	Uganda	none	Qiagen MinElute	200	45
ULIDS-0426	3.04	Uganda	none	Qiagen MinElute	200	45
903577N	3.01	Uganda	none	Qiagen MinElute	200	45
ULIDS-0523	2.99	Uganda	none	Qiagen MinElute	400	45
ULIDS-0171	2.98	Uganda	none	Qiagen MinElute	400	45
ULIDS-0943	2.97	Uganda	none	Qiagen MinElute	400	45
ULIDS-0456	2.96	Uganda	none	Qiagen MinElute	400	45
ULIDS-0595	2.93	Uganda	none	Qiagen MinElute	400	45
ULIDS-0460	2.93	Uganda	none	Qiagen MinElute	400	45
ULIDS-0455	2.92	Uganda	none	Qiagen MinElute	400	45
ULIDS-0955	2.88	Uganda	none	Qiagen MinElute	400	45
ULIDS-0530	2.88	Uganda	none	Qiagen MinElute	400	45
ULIDS-0252	2.68	Uganda	none	Qiagen MinElute	400	45
ULIDS-0600	2.67	Uganda	none	Qiagen MinElute	400	45
906994N	2.66	Uganda	none	Qiagen MinElute	400	45
903576N	2.59	Uganda	none	Qiagen MinElute	75	45
900682N	2.58	Uganda	none	Qiagen MinElute	400	45
ULIDS-0437	2.55	Uganda	none	Qiagen MinElute	200	45
ULIDS-0007	2.54	Uganda	none	Qiagen MinElute	400	45
906813N	2.48	Uganda	none	Qiagen MinElute	200	45
ULIDS-0467	2.48	Uganda	none	Qiagen MinElute	400	45
906051N	2.46	Uganda	none	Qiagen MinElute	400	45
908281N	2.45	Uganda	none	Qiagen MinElute	200	45
906646N	2.44	Uganda	none	Qiagen MinElute	400	45
903602N	2.44	Uganda	none	Qiagen MinElute	200	45
907003N	2.37	Uganda	none	Qiagen MinElute	400	45
ULIDS-0002	2.3	Uganda	none	Qiagen MinElute	400	45
ULIDS-0433	2.3	Uganda	none	Qiagen MinElute	400	45

905960N	2.29	Uganda	none	Qiagen MinElute	400	45
ULIDS-0453	2.24	Uganda	none	Qiagen MinElute	400	45
908241N	2.23	Uganda	none	Qiagen MinElute	400	45
901661N	2.22	Uganda	none	Qiagen MinElute	400	45
ULIDS-0278	2.16	Uganda	none	Qiagen MinElute	400	45
ULIDS-0010	2.14	Uganda	none	Qiagen MinElute	400	45
904800N	2.08	Uganda	none	Qiagen MinElute	400	45
903709N	2.07	Uganda	none	Qiagen MinElute	400	45
907042N	2	Uganda	none	Qiagen MinElute	400	45
904560N	1.97	Uganda	none	Qiagen MinElute	400	45
ULIDS-0431	1.9	Uganda	none	Qiagen MinElute	400	45
906692N	1.89	Uganda	none	Qiagen MinElute	400	45
ULIDS-0440	1.83	Uganda	none	Qiagen MinElute	400	45
ULIDS-0926	1.74	Uganda	none	Qiagen MinElute	400	45
ULIDS-0579	1.72	Uganda	none	Qiagen MinElute	400	45
905476N	1.68	Uganda	none	Qiagen MinElute	400	45
ULIDS-0439	1.62	Uganda	none	Qiagen MinElute	400	45
ULIDS-0597	1.58	Uganda	none	Qiagen MinElute	400	45
ULIDS-0463	1.58	Uganda	none	Qiagen MinElute	400	45
ULIDS-0448	1.57	Uganda	none	Qiagen MinElute	400	45
ULIDS-0055	1.56	Uganda	none	Qiagen MinElute	400	45
ULIDS-0224	1.54	Uganda	none	Qiagen MinElute	400	45
901190N	1.38	Uganda	none	Qiagen MinElute	200	45
ULIDS-0536	1.32	Uganda	none	Qiagen MinElute	400	45

Supplementary Table 8: HBV samples from Kenya

Participant ID	Date	VL	Cohort	Host dep	Extraction	Input vol	Elution vol
SHBV-A3-007	12/09/2023	8.71	Kenya	none	Qiagen MinElute	400	60
SHBV-A4-015	24/04/2024	7.01	Kenya	none	Qiagen MinElute	400	60
SHBV-A4-006	20/03/2024	7.4	Kenya	none	Qiagen MinElute	400	60
SHBV-A4-022	08/05/2024	6.12	Kenya	none	Qiagen MinElute	400	60
SHBV-A1-011	23/08/2023	4.18	Kenya	none	Qiagen MinElute	400	60
SHBV-A3-010	06/12/2023	4.61	Kenya	none	Qiagen MinElute	400	60
SHBV-A4-021	08/05/2024	4.88	Kenya	none	Qiagen MinElute	400	60
SHBV-A4-013	24/04/2024	4.79	Kenya	none	Qiagen MinElute	400	60
SHBV-A4-025	04/06/2024	4.08	Kenya	none	Qiagen MinElute	400	60
SHBV-A3-042	30/04/2024	0 (2.22 repeat)	Kenya	none	Qiagen MinElute	400	60
SHBV-A4-005	28/11/2023	3.8	Kenya	none	Qiagen MinElute	400	60
SHBV-A3-026	07/02/2024	3.48	Kenya	none	Qiagen MinElute	400	60
SHBV-A3-046	14/05/2024	3.49	Kenya	none	Qiagen MinElute	400	60
SHBV-A4-008	20/03/2024	3.68	Kenya	none	Qiagen MinElute	400	60
SHBV-A3-028	21/02/2024	3.78	Kenya	none	Qiagen MinElute	400	60
SHBV-A1-020	07/11/2023	6.05	Kenya	none	Qiagen MinElute	400	60
SHBV-A3-047	15/05/2024	3.25	Kenya	none	Qiagen MinElute	400	60
SHBV-A3-032	12/03/2024	1.65	Kenya	none	Qiagen MinElute	400	60
SHBV-A4-016	30/04/2024	1.51	Kenya	none	Qiagen MinElute	400	60
SHBV-A4-014	24/04/2024	2.92	Kenya	none	Qiagen MinElute	400	60

Appendix 3: Supplementary material for Chapter 5

Supplementary text 1: Search strategy for HBVDR systematic review and meta-analysis

PubMed

(((((("Hepatitis B virus"[Mesh]) OR ("hepatitis b"[Title/Abstract] OR CHB[Title/Abstract] OR HPV[Title/Abstract])) AND ((("Entecavir" [Supplementary Concept]) OR "Tenofovir"[Mesh]) OR (Entecavir[Title/Abstract] OR Tenofovir[Title/Abstract] OR Viread[Title/Abstract])))) AND (("Drug Resistance"[Mesh]) OR (resistan*[Title/Abstract] OR "drug mutat*"[Title/Abstract] OR DRMs[Title/Abstract] OR escap*[Title/Abstract] OR mutat*[Title/Abstract]))) AND (chronic*)

Filters applied: English.

Database: Embase 1974 to present

Link to search history: https://ovidsp.ovid.com/ovidweb.cgi?T=JS&NEWS=N&PAGE=main&SHAREDSEARCHID=4n05jJsL1ji4FTF87qkNYBsZHPU9nRUTS2dvJ0VY0eGxMFI8LbocaxjpUQAhZZhHX

Search Strategy:

1 exp Hepatitis B virus/ (62830)

2 ("hepatitis b" or CHB or HBV).ti,ab. (149344)

3 1 or 2 (158457)

4 Entecavir/ (11052)

5 Tenofovir/ (21517)

6 (Entecavir or Tenofovir or Viread).ti,ab. (20044)

7 4 or 5 or 6 (35130)

8 exp drug resistance/ (374437)

9 (resistan* or "drug mutat*" or DRMs or escap* or mutat*).ti,ab. (2570076)

10 8 or 9 (2642215)

11 exp chronic viral hepatitis/ (32355)

12 chronic*.ti,ab,kw. (2067571)

13 11 or 12 (2072437)

14 3 and 7 and 10 and 13 (2870)

15 limit 14 to english language (2649)

Database: Global Health <1973 to 2023 Week 21>

Link to search history: https://ovidsp.ovid.com/ovidweb.cgi?T=JS&NEWS=N&PAGE=main&SHAREDSEARCHID=4oiBkIUlltcJfuJ4tgWl6jFpYipaXpXfRQ03vmUTgnLPITNDPCMrv6Kg3xNZrX6Kp

Search Strategy:

1 exp hepatitis b virus/ (39525)

2 ("hepatitis b" or CHB or HBV).ti,ab. (42432)

3 1 or 2 (43954)

4 Entecavir/ (1174)

5 Tenofovir/ (3980)

6 (Entecavir or Tenofovir or Viread).ti,ab. (5825)

7 4 or 5 or 6 (6014)

8 exp drug resistance/ (139291)

9 (resistan* or "drug mutat*" or DRMs or escap* or mutat*).ti,ab. (347385)

10 8 or 9 (355161)

11 chronic infections/ (23637)

12 chronic*.ti,ab. (222202)

13 11 or 12 (222679)

14 3 and 7 and 10 and 13 (615)

15 limit 14 to english language (539)

Scopus

(TITLE-ABS-KEY ("hepatitis b" OR chb OR hbv) AND TITLE-ABS-KEY (Entecavir OR Tenofovir OR viread) AND TITLE-ABS-KEY (resistan* OR "drug mutat*" OR drms OR escap* OR mutat*) AND TITLE-ABS-KEY (chronic*)) AND (LIMIT-TO (LANGUAGE , "English"))

Cochrane Central Register of Controlled Trials

Issue 6 of 12, June 2023

#1 ("hepatitis b" or CHB or HBV):ti,ab,kw 10780

#2 (Entecavir or Tenofovir or Viread):ti,ab,kw 4649

#3 (resistan* or "drug mutat*" or DRMs or escap* or mutat*):ti,ab,kw 103427

#4 chronic*:ti,ab,kw 178032

#5 #1 and #2 and #3 and #4 425

Clinicaltrials.gov

Condition or disease: Hepatitis B, Chronic, Other terms: Entecavir

Condition or disease: Hepatitis B, Chronic, Other terms: Tenofovir

ISRTCN Registry

Condition: chronic hepatitis b, Interventions: Entecavir
https://www.isrctn.com/search?q=&filters=condition%3Achronic+hepatitis+b%2Cintervention%3AEntecavir

Condition: chronic hepatitis b, Interventions: Tenofovir
https://www.isrctn.com/search?q=&filters=condition%3Achronic+hepatitis+b%2Cintervention%3ATenofovir

WHO International Clinical Trials Registry Platform
https://trialsearch.who.int/AdvSearch.aspx

Title: resistance, Condition: chronic hepatitis b, Intervention: Entecavir, ALL trials

Title: resistance, Condition: chronic hepatitis b, Intervention: Tenofovir, ALL trials

Supplementary table 9: PRISMA checklist for systematic review and meta-analysis of tenofovir and entecavir resistance[299]

Section and Topic	Item #	Checklist item	Location where item is reported
TITLE			
Title	1	Identify the report as a systematic review.	Title
ABSTRACT			
Abstract	2	See the PRISMA 2020 for Abstracts checklist.	Abstract
INTRODUCTION			
Rationale	3	Describe the rationale for the review in the context of existing knowledge.	Intro
Objectives	4	Provide an explicit statement of the objective(s) or question(s) the review addresses.	Intro
METHODS			
Eligibility criteria	5	Specify the inclusion and exclusion criteria for the review and how studies were grouped for the syntheses.	Methods
Information sources	6	Specify all databases, registers, websites, organisations, reference lists and other sources searched or consulted to identify studies. Specify the date when each source was last searched or consulted.	Supplement
Search strategy	7	Present the full search strategies for all databases, registers and websites, including any filters and limits used.	Supplement
Selection process	8	Specify the methods used to decide whether a study met the inclusion criteria of the review, including how many reviewers screened each record and each report retrieved, whether they worked independently, and if applicable, details of automation tools used in the process.	Supplement
Data collection process	9	Specify the methods used to collect data from reports, including how many reviewers collected data from each report, whether they worked independently, any processes for obtaining or confirming data from study investigators, and if applicable, details of automation tools used in the process.	Methods/Supplement
Data items	10a	List and define all outcomes for which data were sought. Specify whether all results that were compatible with each outcome domain in each study were sought (e.g. for all measures, time points, analyses), and if not, the methods used to decide which results to collect.	Methods/Supplement
	10b	List and define all other variables for which data were sought (e.g. participant and intervention characteristics, funding sources). Describe any assumptions made about any missing or unclear information.	Supplement
Study risk of bias assessment	11	Specify the methods used to assess risk of bias in the included studies, including details of the tool(s) used, how many reviewers assessed each study and whether they worked independently, and if applicable, details of automation tools used in the process.	Methods/Supplement
Effect measures	12	Specify for each outcome the effect measure(s) (e.g. risk ratio, mean difference) used in the synthesis or presentation of results.	Methods

Synthesis methods	13a	Describe the processes used to decide which studies were eligible for each synthesis (e.g. tabulating the study intervention characteristics and comparing against the planned groups for each synthesis (item #5)).	Methods
	13b	Describe any methods required to prepare the data for presentation or synthesis, such as handling of missing summary statistics, or data conversions.	Methods
	13c	Describe any methods used to tabulate or visually display results of individual studies and syntheses.	Methods
	13d	Describe any methods used to synthesize results and provide a rationale for the choice(s). If meta-analysis was performed, describe the model(s), method(s) to identify the presence and extent of statistical heterogeneity, and software package(s) used.	Methods
	13e	Describe any methods used to explore possible causes of heterogeneity among study results (e.g. subgroup analysis, meta-regression).	Methods
	13f	Describe any sensitivity analyses conducted to assess robustness of the synthesized results.	Methods
Reporting bias assessment	14	Describe any methods used to assess risk of bias due to missing results in a synthesis (arising from reporting biases).	Methods
Certainty assessment	15	Describe any methods used to assess certainty (or confidence) in the body of evidence for an outcome.	Methods
RESULTS			
Study selection	16a	Describe the results of the search and selection process, from the number of records identified in the search to the number of studies included in the review, ideally using a flow diagram.	Figure 2
	16b	Cite studies that might appear to meet the inclusion criteria, but which were excluded, and explain why they were excluded.	Discussion
Study characteristics	17	Cite each included study and present its characteristics.	Table 5
Risk of bias in studies	18	Present assessments of risk of bias for each included study.	Table 5
Results of individual studies	19	For all outcomes, present, for each study: (a) summary statistics for each group (where appropriate) and (b) an effect estimate and its precision (e.g. confidence/credible interval), ideally using structured tables or plots.	Results figures
Results of syntheses	20a	For each synthesis, briefly summarise the characteristics and risk of bias among contributing studies.	Results
	20b	Present results of all statistical syntheses conducted. If meta-analysis was done, present for each the summary estimate and its precision (e.g. confidence/credible interval) and measures of statistical heterogeneity. If comparing groups, describe the direction of the effect.	Results
	20c	Present results of all investigations of possible causes of heterogeneity among study results.	Results t
	20d	Present results of all sensitivity analyses conducted to assess the robustness of the synthesized results.	Resultst
Reporting biases	21	Present assessments of risk of bias due to missing results (arising from reporting biases) for each synthesis assessed.	Results
Certainty of evidence	22	Present assessments of certainty (or confidence) in the body of evidence for each outcome assessed.	Results
DISCUSSION			

Discussion	23a	Provide a general interpretation of the results in the context of other evidence.	Discussion
	23b	Discuss any limitations of the evidence included in the review.	Discussion
	23c	Discuss any limitations of the review processes used.	Discussion
	23d	Discuss implications of the results for practice, policy, and future research.	Discussion
OTHER INFORMATION			
Registration and protocol	24a	Provide registration information for the review, including register name and registration number, or state that the review was not registered.	Methods
	24b	Indicate where the review protocol can be accessed, or state that a protocol was not prepared.	Methods
	24c	Describe and explain any amendments to information provided at registration or in the protocol.	Methods
Support	25	Describe sources of financial or non-financial support for the review, and the role of the funders or sponsors in the review.	Funding
Competing interests	26	Declare any competing interests of review authors.	Competing interests
Availability of data, code and other materials	27	Report which of the following are publicly available and where they can be found: template data collection forms; data extracted from included studies; data used for all analyses; analytic code; any other materials used in the review.	Data via supplementary file

*Figure numbering corresponds to that in published manuscript

Supplementary table 10: Analysis of resistance criteria for TFV papers included in HBVDR systematic review and meta-analysis

Study ID	'Established' TFV RAMs in EASL guideline (A181T/V and N236T) (Table 1A)		'Putative' RAM combinations (Table 1B)	
	Identified	Classified as resistant	Identified	Classified as resistant
Both established and/or putative RAMs classified as resistant				
Bakhshizadeh 2015	N	Y	N	Y (A194T)
Hongthanakorn 2011	N	Y	N	Y (A194T)
Patterson 2011	Y	Y	NS	NS
Tan 2008	Y	Y	NS	NS
Liang 2019	N	Y	L180M+M204I/V +V173L	N
DeFrancesco 2015	N	N	1)A194T, 2)L180M+M204I/V +V173L	1) Y (A194T), 2) N
Neither established nor putative RAMs classified as resistant				
Corsa 2014	Y	N	A194A/P	N
Hou 2015	Y	N	(V173L/V/L180M/L/M204V/M/S213T)	N
Boyd 2014	N	N (Geno2pheno*)	L180M+M204I/V +V173L	N (Geno2pheno*)
Snow-Lampart 2011	N	N	L180M+M204V +V173L	N
Stephan 2005	NS	N	L180M+M204V +V173L	N
Boyd 2015	N	NS	L180M+M204I/V +V173L	N
Singla 2015	N	N (Geno2pheno*)	N	N (Geno2pheno*)
Established RAMs not classified as resistant, putative not reported				
Cathcart 2018	Y	N	N	NS
Lim 2016 (2)	Y (all A181V/T and/or N236T at baseline)	N	N	NS
Yang 2015	Y at baseline	N	NS	NS
Berg 2014	Y	N	NS	NS
Liu 2017	Y	N	NS	NS
RAM criteria not specified, RAM details not reported				
Lim 2016 (1)	Y (individuals excluded if identified at baseline)	NS	N	NS
Fung 2017	N	NS	N	NS
Marcellin 2008	NS	NS	NS	NS
Pan 2014	NS	NS	NS	NS
Chan 2023	NS	NS	NS	NS
Srivastava 2016	NS	NS	NS	NS
Buti 2015	NS	NS	NS	NS
Cho 2015	NS	NS	NS	NS

*NB HBV geno2pheno does not list any TFV RAMs

Supplementary table 11: Characteristics of 62 studies included in the systematic review. Where multiple subgroups of individuals are reported in the study, only subgroups meeting inclusion criteria are detailed in the table.

Study				Intervention		Population					Outcome		Bias
Study	Study design	WHO region	Study size and group	Drug/ Dose	Duration (m)	Prior NA exposure	Baseline GR*	Age group	HBeAg status	HIV co-infection	Seq. criteria	GR risk (%)	Risk of bias (1 = high, 5 = low)
Bakhshizadeh 2015[300]	Cohort/cross sectional; Prosp.	EMR	93 NT	TDF 300mg OD	mean 21.10 (R 6-36)	Naïve	NA	Adults	eAg+ and eAg-	No	All	0.0	5
Berg 2014[301]	Clinical trial; Prospective	AMR, EMR	53 ET	TDF 300mg OD	EP 38.00	Experienced	All	Adults	eAg+ and eAg-	Not reported	VBT only	0.0	2
Boyd 2014[302]	Cohort/cross sectional; Prosp.	EUR	24 ET	TDF (as ART)	med 74.70 (IQR 33.4-94.7)	Experienced	Not reported	Adults	eAg+ and eAg-	Yes	Unclear	0.0	4
Boyd 2015[303]	Cohort/cross sectional; Prosp.	AFR	86 NT	TDF (as ART)	med 24.30 (IQR 17.6-30.8)	Naïve	NA	Adults	eAg+ and eAg-	Yes	All	0.0	5
Buti 2015[304]	Clinical trial; Prospective	AMR, EUR, WPR	585 NT	TDF 300mg OD	EP 77.3	Naïve	NA	Adults	eAg+ and eAg-	Not reported	All	0.0	3
Cathcart 2018[305]	Clinical trial; Prospective	AMR, SEAR, EUR, WPR	1002 NT 296 ET	TAF 25mg, TDF 300mg	EP 22.09	Naïve and experienced	Mix	Adults	eAg+ and eAg-	No	All	0.0 NT 0.0 ET	5
Chan 2023[103]	Clinical trial; Prospective	AMR, SEAR, EUR, EMR, WPR	993 NT 305 ET	TAF 25mg, TDF 300mg	EP 55.2	Naïve and experienced	Not reported	Adults	eAg+ and eAg-	No	All	0.0 NT 0.0 ET	4
Chang 2005[306]	Clinical trial; Prospective	AMR, EUR, EMR, WPR	39 EE	ETV 1mg OD	EP 11.04	Experienced	All	Adults	eAg+ and eAg-	No	All	0.0	4
Chang 2006[250]	Clinical trial; Prospective	AMR, SEAR, EUR, WPR	344 NE 10 EE	ETV 0.5 mg OD	EP 11.96	Naïve and experienced	NA	Adults	eAg+ and eAg-	No	All	0.0 NE 0.0 EE	4
Chang 2009[307]	Clinical trial; Prospective	AMR, SEAR, EUR, WPR	243 NE	ETV 0.5 mg OD	EP 24.00	Naïve	NA	Adults	all eAg+	No	All	0.0	4
Chen 2011[308]	Cohort/cross sectional; Prosp.	WPR	48 NE	ETV 0.5 mg OD	EP 24.00	Naïve	NA	Adults	eAg+ and eAg-	No	VBT only	0.0	5
Cho 2015[247]	Cohort or cross sectional; Retros	WPR	146 EE 56 ET	ETV 0.5/1mg OD. TDF 300mg OD	med 37.70 (IQR 23.4-74.5)	Experienced	No	Adults	eAg+ and eAg-	No	All	8.9 EE 0.0 ET	4
Cho 2017[309]	Cohort/cross sectional; Prosp.	WPR	1009 NE	ETV 0.5 mg OD	med 26.50 (R 6-77.4)	Naïve	NA	Adults	eAg+ and eAg-	No	Unclear	1.2	4
Corsa 2014[310]	Clinical trial; Prospective	AMR, EUR, WPR	280 ET	TDF 300mg OD	EP 22.09	Experienced	All	Adults	eAg+ and eAg-	No	All	0.0	5
DeFrancesco 2015[244]	Cohort or cross sectional; Retrospective	EUR	11 ET	TDF 300mg OD	EP 24.00	Experienced	All	Adults	eAg+ and eAg-	Not reported	VBT only	9.1	3
Deng 2013[311]	Cohort/cross sectional; Prosp.	WPR	21 EE	ETV 1mg OD	EP 33.14	Experienced	All	Adults	all eAg+	Not reported	All	23.4	3
Fung 2014[312]	Clinical trial; Prospective	AMR, EUR, WPR	141 ET	TDF 300mg OD	EP 22.09	Experienced	All	Adults	eAg+ and eAg-	No	All	0.0	5
Fung 2017[102]	Clinical trial; Prospective	AMR, EUR, WPR	141 ET	TDF 300mg OD	EP 55.23	Experienced	All	Adults	eAg+ and eAg-	No	All	0.0	5
Gish 2007[313]	Clinical trial; Prospective	AMR	243 NE	ETV 0.5 mg OD	EP 22.09	Naïve	NA	Adults	all eAg+	Not reported	All	0.0	5
Gwak 2013[314]	Cohort or cross sectional; Prospective	WPR	58 NE	ETV 0.5 mg OD	EP 24.00	Naïve	NA	Adults	eAg+ and eAg-	No	VBT only	0.0	5

Ha 2011[315]	Cohort or cross sectional; Retros	AMR	107 NE	ETV 0.5 mg OD	EP 48.00	Naïve	NA	Adults	all eAg-	No	VBT only	0.0	5
Heathcote 2011[316]	Clinical trial; Prospective	AMR, EUR, WPR	196 ET	TDF 300mg OD	EP 22.09	Experienced	No	Adults	eAg+ and eAg-	No	All	0.0	5
Heo 2012[246]	Clinical trial; Prospective	WPR	34 EE	ETV 1mg OD	EP 22.09	Experienced	No	Adults	all eAg+	No	All	2.9	5
Hongthanakorn 2011[89]	Cohort or cross sectional; Retros	AMR	15 ET 43 NE 13 EE	ETV, TDF, dose NS	mean 37.50 ET 37.50 NE 37.5 EE (R 12-102)	Naïve and experienced	Not reported	Adults	eAg+ and eAg-	No	VBT only	0.0 ET 2.3 NE 38.5 EE	4
Hou 2015[241]	Clinical trial; Prospective	WPR	244 NT 11 ET	TDF 300mg OD	EP 11.04	Naïve and experienced	Mix	Adults	eAg+ and eAg-	No	All	0.4 NT 0.0 ET	4
Jonas 2016[92]	Clinical trial; Prospective	AMR, SEAR, EUR	120 NE	ETV 0.015 mg/kg OD, max 0.5 mg	22.09	Naïve	NA	Children	all eAg+	No	All	3.3	4
Kamezaki 2011[317]	Cohort or cross sectional; Retro	WPR	81 NE	ETV 0.5 mg OD	mean 27.00	Naïve	NA	Adults	eAg+ and eAg-	No	VBT only	2.5	5
Kamezaki 2013[318]	Cohort or cross sectional; Retro	WPR	135 NE	ETV 0.5 mg OD	26.90 (SD +/- 321.6)	Naïve	NA	Adults	eAg+ and eAg-	No	Unclear	1.5	5
Karino 2010[319]	Cohort or cross sectional; Prosp	WPR	42 EE	ETV 1mg OD	EP 34.06	Experienced	All	Adults	eAg+ and eAg-	No	All	30.9	3
Kim 2010 (1)[320]	Cohort or cross sectional; Retro	WPR	73 NE	ETV 0.5 mg OD	mean 18.40 (SD +/- 3.8)	Naïve	NA	Adults	eAg+ and eAg-	No	VBT only	0.0	4
Kim 2010 (2)[321]	Cohort or cross sectional; Retro	WPR	24 EE	ETV 1mg OD	EP 24.00	Experienced	All	Adults	eAg+ and eAg-	No	VBT only	25.0	5
Kim 2017[93]	Cohort or cross sectional; Retro	WPR	202 NE 56 EE	ETV 0.5 mg OD	median 59 NE/ EE (R9-101)	Naïve and experienced	No	Adults	eAg+ and eAg-	No	VBT only	3.0 NE 3.6 EE	4
Kitrinos 2014[273]	Clinical trial; Prospective	AMR, EUR, WPR	389 ET	TDF 300mg OD	EP 72.00	Experienced	Mix	Adults	eAg+ and eAg-	No	All	0.0	3
Liang 2019[322]	Clinical trial; Prospective	WPR	257 NT	TDF 300mg OD	EP 55.00	Naive	NA	Adults	eAg+ and eAg-	No	All	0.0	5
Lim 2016 (1)[323]	Clinical trial; Prospective	WPR	44 ET	TDF 300mg OD	EP 11.09	Experienced	All	Adults	eAg+ and eAg-	No	All	0.0	5
Lim 2016 (2)[324]	Clinical trial; Prospective	WPR	50 ET	TDF 300mg OD	EP 22.09	Experienced	All	Adults	eAg+ and eAg-	No	All	0.0	5
Liu 2016[99]	Cohort/cross sectional; Prosp.	WPR	33 NE 56 EE	ETV 0.5mg naive, 1mg experienced	69.00 NE (R 60-75), 57.00 EE (R12-75)	Naïve and experienced	Mix	Adults	eAg+ and eAg-	No	VBT only	0.0 NE 16.1 EE	5
Liu 2017[325]	Clinical trial; Prospective	AMR, EUR, WPR	426 ET	TDF 300mg OD	EP 96.00	Experienced	Mix	Adults	eAg+ and eAg-	No	All	0.0	4
Lok 2012[76]	Clinical trial; Prospective	ARF, AMR, EUR, SEAR, WPR	182NE	ETV 0.5 mg OD	EP 23.01	Naïve	NA	Adults; Children	eAg+ and eAg-	Not reported	All	0.0	3
Marcellin 2008[326]	Clinical trial; Prospective	AMR, EUR, WPR	375 NT 51 ET	TDF 300mg OD	EP 11.04	Naïve and experienced	Not reported	Adults	eAg+ and eAg-	No	All	0.0 NT 0.0 ET	4
Pan 2014[327]	Clinical trial; Prospective	AMR	87 NT	TDF 300mg OD	EP 11.04	Naïve	NA	Adults	eAg+ and eAg-	Not reported	All	0.0	5
Park 2011[88]	Cohort or cross sectional; Prospective	WPR	55 EE	ETV 1mg OD	median 24.0 (12-47)	Experienced	All	Adults	eAg+ and eAg-	No	VBT only	36.4	5

Patterson 2011[242]	Clinical trial; Prospective	WPR	38 ET	TDF 300mg OD	EP 22.09	Experienced	All	Adults	eAg+ and eAg-	No	VBT only	7.9	5
Sherman 2006[328]	Clinical trial; Prospective	AMR, SEAR, EUR, EMR, WPR	141 ET	ETV 1mg OD	EP 11.05	Experienced	All	Adults	all eAg+	No	All	1.4	4
Shin 2011[329]	Cohort or cross sectional; Retro	WPR	61 NE	ETV 0.5mg OD	EP 11.05	Naïve	NA	Adults	eAg+ and eAg-	No	VBT only	0.0 NE	4
Singla 2015[330]	Cohort/cross sectional; Prosp.	SEAR	30 NT 39 NE	TDF 300mg OD ETV 0.5mg OD	EP 12.00	Naïve	NA	Adults	eAg+ and eAg-	Not reported	All	0.0 NT 0.0 NE	3
Snow-Lam part 2011[331]	Clinical trial; Prospective	AMR, EUR, WPR	196 TE	TDF 300mg OD	EP 22.09	Experienced	Mix	Adults	eAg+ and eAg-	No	All	0.0	4
Srivastava 2016[245]	Cohort/cross sectional; Prosp.	SEAR	26 NT 25 NE	ETV 0.5mg OD TDF 300mg OD	EP 24.00	Naïve	NA	Adults; Children	eAg+ and eAg-	No	All	0.0 NT 0.0 NE	4
Stephan 2005[332]	Cohort or cross sectional; Retro	EUR	24 ET	TDF (as ART)	EP 11.05	Experienced	Not reported	Adults	eAg+ and eAg-	Yes	All	0.0	5
Suzuki 2008[333]	Clinical trial; Prospective	WPR	84 EE	ETV 0.5mg, 1mg OD	EP 12.00	Experienced	All	Adults	eAg+ and eAg-	No	VBT only	0.0	3
Suzuki 2010[249]	Cohort or cross sectional; Retro	WPR	130 EE	ETV 0.5 mg OD	EP 12.00	Experienced	All	Adults	eAg+ and eAg-	Not reported	Unclear	0.8	5
Suzuki 2019[334]	Cohort or cross sectional; Retro	WPR	1094 NE	ETV 0.5mg	median 66.00 (R 12 - 120)	Naïve	NA	Adults	eAg+ and eAg-	No	VBT only	0.7	5
Tan 2008[243]	Cohort or cross sectional; Retro	AMR	10 ET	TDF 300mg OD	EP 20.50	Experienced	Mix	Adults	eAg+ and eAg-	Not reported	All	10.0	3
Tenney 2007[335]	Cohort/cross sectional; Prosp.	AMR, EUR, WPR	192 EE	ETV 1mg OD	EP 24.00	Experienced	All	Adults	eAg+ and eAg-	Not reported	All	12.0	2
Tenney 2009[91]	Clinical trial; Prospective	AMR, SEAR, EUR, WPR	108 NE 33 EE	ETV 0.5mg naive, 1mg exp	EP 60.00	Naïve and experienced	All	Adults	eAg+ and eAg-	Not reported	All	8.3 NE 42.4 EE	4
Tsai 2012[336]	Cohort/cross sectional; Prosp.	WPR	98 NE	ETV 0.5 mg OD	EP 11.05	Naïve	NA	Adults	eAg+ and eAg-	Not reported	VBT only	0.0 NE	4
Wang 2013[248]	Cohort/cross sectional; Prosp.	WPR	25 EE	ETV 0.5 mg OD	EP 12.00	Experienced	No	Adults	all eAg+	No	All	4.0	3
Xu 2022[90]	Clinical trial; Prospective	WPR	208 NE 18 EE	ETV 0.5mg OD	EP 44.00	Naïve and experienced	Mix	Adults	eAg+ and eAg-	Not reported	All	1.4 NE 22.2 EE	5
Yang 2015[337]	Cohort or cross sectional; Retro	WPR	28 ET	TDF 300mg OD	EP 11.05	Experienced	All	Adults	eAg+ and eAg-	No	All	0.0 ET	5
Yokosuka 2010[338]	Clinical trial; Prospective	WPR	66 NE	ETV 0.01, 0.1, 0.5mg then 0.5mg rollover	EP 27.6	Naïve	NA	Adults	eAg+ and eAg-	No	All	1.5	5
Yuen 2015[339]	Clinical trial; Prospective	WPR	30 NE	ETV 0.5 mg OD	EP 22.09	Naïve	NA	Adults	eAg+ and eAg-	No	All	0.0	5
Zhou 2017[340]	Cohort/cross sectional; Prosp.	WPR	33 ET	TDF 300mg OD	EP 11.05	Experienced	All	Adults	eAg+ and eAg-	No	Unclear	0.0 ET	5

Abbreviations: ART = antiretroviral therapy, eAg- = HBeAg negative, eAg+ = HBeAg positive, EE = NA experienced on Entecavir, ET = NA experienced on tenofovir, GR = genotypic resistance, NA = nucleoside analogue, NE = NA naive on Entecavir, NT = NA naive on tenofovir, OD = once a day, Prosp = prospective, Retro = retrospective, TAF = tenofovir alafenamide, TDF = tenofovir disoproxil. VBT = virological breakthrough, * RAM identified for any NA

Supplementary table 12: Sensitivity analysis for Experienced/Tenofovir (n=19)

** Or single study estimate if only 1 study remaining for analysis

Year group	Number of studies remaining	Pooled risk estimate from random effects model*	95% CI lower	95% CI upper	Heterogeneity (I^2)
Baseline analysis					
2	8	0.000	0.000	0.006	66%
Study size >30					
2	6	0.000	0.000	0.005	56%
Exclude studies with high risk of bias					
2	8	0.000	0.000	0.006	66%
Study design = clinical trial					
2	6	0.000	0.000	0.005	56%
Sequencing criteria = All sequenced					
2	6	0.000	0.000	0.000	4%
Baseline mutations - exclude those where all individuals have RAMs at baseline					
2	4	0.000	0.000	0.000	39%

Supplementary Table 13: Sensitivity analysis for Naive/Entecavir (n = 22)

Year group	Number of studies remaining	Pooled risk estimate from random effects model*	95% CI lower	95% CI upper	Heterogeneity (I^2)
Baseline analysis					
2	13	0.003	0.000	0.009	41%
4	2	0.006	0.000	0.027	47%
≥5	4	0.009	0.001	0.023	49%
Study size >30					
2	11	0.004	0.000	0.011	51%
4	2	0.006	0.000	0.027	47%
≥5	4	0.009	0.001	0.023	49%
Study design = clinical trial					
2	6	0.002	0.000	0.013	60%
4	1	0.014	0.003	0.042	NA
≥5	1	0.009	0.000	0.051	NA
Sequencing criteria = All sequenced					
2	7	0.001	0.000	0.010	53%
4	1	0.014	0.003	0.042	NA
≥5	1	0.009	0.000	0.051	NA
Baseline mutations - exclude those where all individuals have RAMs at baseline					
2	13	0.003	0.000	0.009	41%
4	2	0.006	0.000	0.027	47%
≥5	3	0.009	0.000	0.031	65%

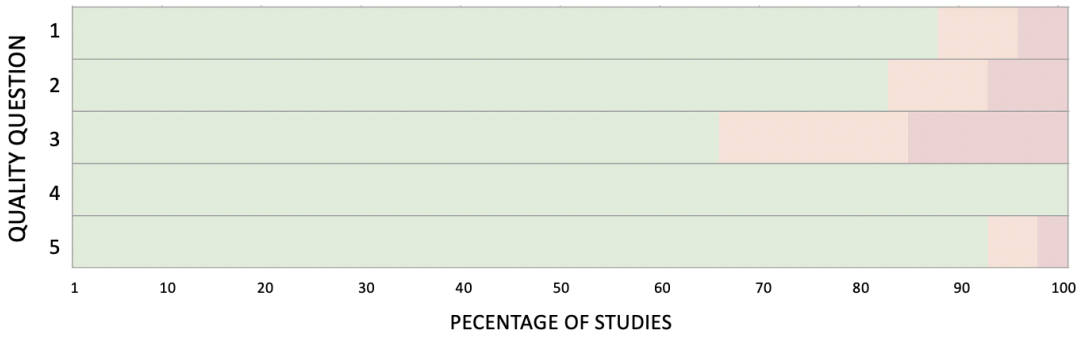
* Treatment duration (years)

** Or single study estimate if only 1 study remaining for analysis

Supplementary Table 14: Sensitivity analysis for Experienced/Entecavir (n = 18)

Year group	Number of studies remaining	Pooled risk estimate from random effects model*	95% CI lower	95% CI upper	Heterogeneity (I^2)
Baseline analysis					
2	4	0.170	0.054	0.329	87%
3	4	0.226	0.082	0.410	82%
≥5	2	0.201	0.016	0.501	93%
Study size >30					
2	3	0.151	0.026	0.342	90%
3	2	0.180	0.020	0.439	91%
≥5	3	0.201	0.016	0.501	93%
Exclude studies with high risk of bias					
2	3	0.193	0.023	0.454	88%
3	4	0.226	0.082	0.410	82%
≥5	3	0.201	0.016	0.501	93%
Sequencing criteria = All sequenced					
2	2	0.181	0.017	0.179	62%
3	3	0.193	0.064	0.476	84%
≥5	1	0.515	0.335	0.692	NA
Baseline mutations - exclude those where all individuals have RAMs at baseline					
2	1	0.029	0.001	0.153	NA
3	2	0.193	0.000	0.545	85%
≥5	2	0.089	0.005	0.244	80%

* Treatment duration (years), ** Or single study estimate if only 1 study remaining for analysis



Supplementary Figure 1: Summary of risk of bias scores for included studies.

Percentage of studies rated as “yes” (green), “unclear” (orange) or “no” (red) to five quality Joanna Briggs Institute bias scoring questions:

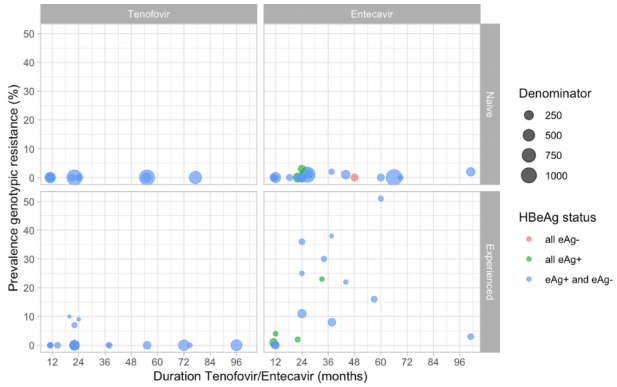
Q1. Were the criteria for inclusion in the study clearly defined?

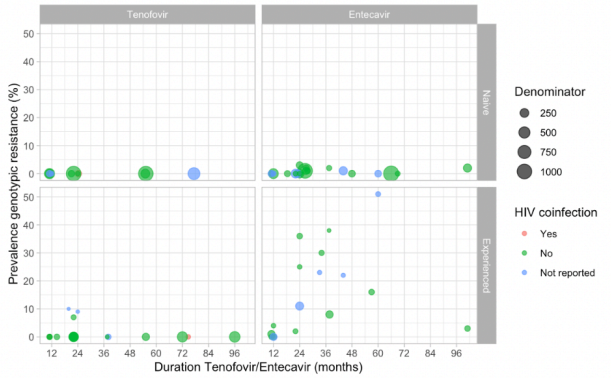
*Q2. Were the study subjects and the se**tting described in detail?***

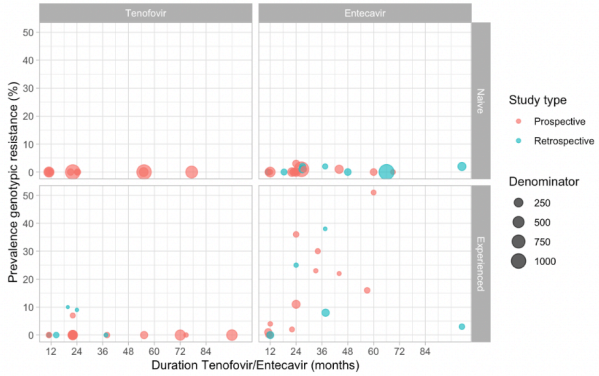
Q3. Was the exposure (drug treatment) measured in a valid and reliable way?

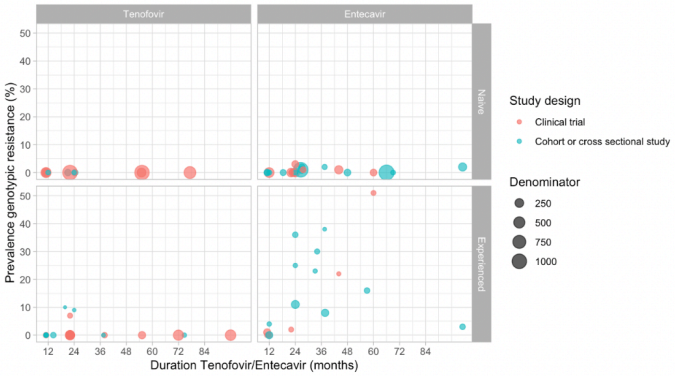
Q4. Were objective, standard criteria used for measurement of the condition (ie. chronic HBV)?

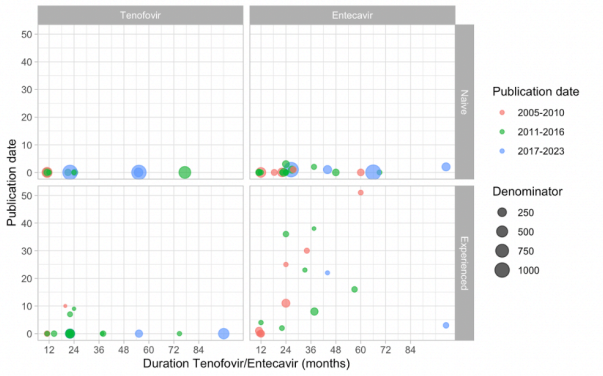
Q5. Were the outcomes measured in a valid and reliable way?

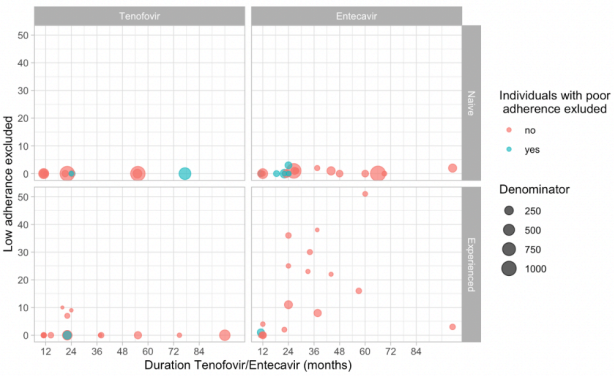

Tenofovir
Entecavir
Naive
Experienced
Prevalence genotypic resistance (%)
Duration Tenofovir/Entecavir (months)
Denominator
250
500
750
1000
HBeAg status
all eAg-
all eAg+
eAg+ and eAg-

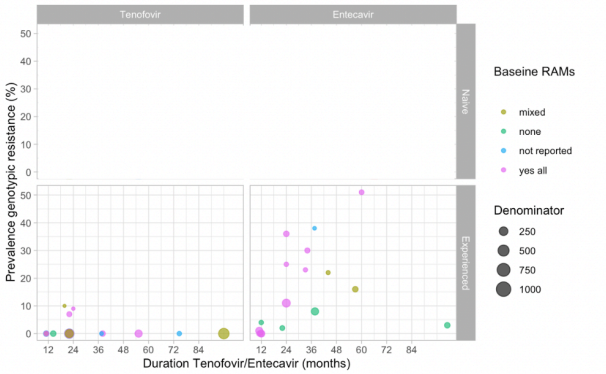

Tenofovir
Entecavir
Naive
Experienced
Prevalence genotypic resistance (%)
Duration Tenofovir/Entecavir (months)
Denominator
250
500
750
1000
HIV coinfection
Yes
No
Not reported

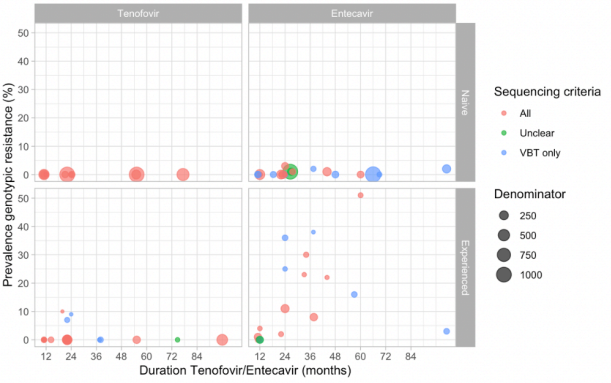

Tenofovir
Entecavir
Naive
Experienced
Prevalence genotypic resistance (%)
Duration Tenofovir/Entecavir (months)
Study type
Prospective
Retrospective
Denominator
250
500
750
1000


Tenofovir
Entecavir
Naive
Experienced
Prevalence genotypic resistance (%)
Duration Tenofovir/Entecavir (months)
Study design
Clinical trial
Cohort or cross sectional study
Denominator
250
500
750
1000

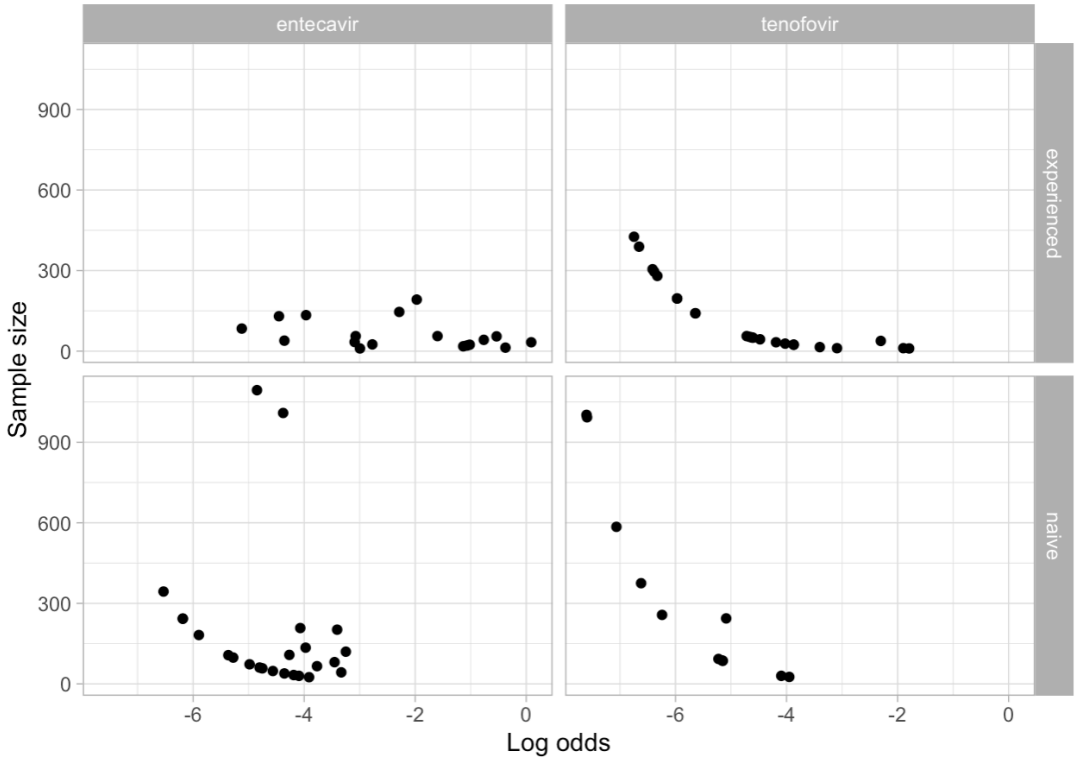

Tenofovir
Entecavir
Naive
Experienced
Publication date
Duration Tenofovir/Entecavir (months)
Publication date
2005-2010
2011-2016
2017-2023
Denominator
250
500
750
1000


Tenofovir
Entecavir
Naive
Experienced
Low adherance excluded
Duration Tenofovir/Entecavir (months)
Individuals with poor adherence exluded
no
yes
Denominator
250
500
750
1000


Tenofovir
Entecavir
Naive
Experienced
Prevalence genotypic resistance (%)
Duration Tenofovir/Entecavir (months)
Baseine RAMs
mixed
none
not reported
yes all
Denominator
250
500
750
1000


Tenofovir
Entecavir
Naive
Experienced
Prevalence genotypic resistance (%)
Duration Tenofovir/Entecavir (months)
Sequencing criteria
All
Unclear
VBT only
Denominator
250
500
750
1000

Supplementary Figure 2: Relationship between clinical and methodological sources of heterogeneity (publication date, HBeAg status, HIV status, study type, publication date, exclusion of those with adherence issues, baseline resistance status and sequencing criteria) and the observed risk of genotypic resistance over time. All 62 studies are shown without deduplication of cohorts where the same data is published multiple times.



Supplementary Figure 3: Funnel plots to investigate publication bias

References

1. Maccallum, F. O. & Bauer, D. J. Homologous serum jaundice transmission experiments with human volunteers. *Lancet* **243**, 622–627 (1944).
2. CDC. CDC Museum COVID-19 Timeline. *Centers for Disease Control and Prevention* https://www.cdc.gov/museum/timeline/covid19.html (2023).
3. Zhu, N. *et al.* A novel Coronavirus from patients with pneumonia in China, 2019. *N. Engl. J. Med.* **382**, 727–733 (2020).
4. Aspinall, E. COVID-19 Timeline. *British Foreign Policy Group* https://bfpg.co.uk/2020/04/covid-19-timeline/#:~:text=The%20UK's%20first%20two%20patients,arrives%20at%20RAF%20Brize%20Norton. (2020).
5. COVID-19 deaths. *datadot* https://data.who.int/dashboards/covid19/deaths.
6. COVID-19 cases. *datadot* https://data.who.int/dashboards/covid19/cases.
7. Zhang, Q. *et al.* Molecular mechanism of interaction between SARS-CoV-2 and host cells and interventional therapy. *Signal Transduct. Target. Ther.* **6**, 233 (2021).
8. Hodcroft, E. B. *et al.* Spread of a SARS-CoV-2 variant through Europe in the summer of 2020. *Nature* **595**, 707–712 (2021).
9. Rambaut, A. *et al.* A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. *Nat. Microbiol.* **5**, 1403–1407 (2020).
10. auspice. https://nextstrain.org/ncov/gisaid/global/all-time.
11. Coronavirus disease (COVID-19). https://www.who.int/health-topics/coronavirus#tab=tab_1.
12. Davis, H. E., McCorkell, L., Vogel, J. M. & Topol, E. J. Long COVID: major findings, mechanisms and recommendations. *Nat. Rev. Microbiol.* **21**, 133–146 (2023).
13. Jayk Bernal, A. *et al.* Molnupiravir for oral treatment of Covid-19 in nonhospitalized patients. *N. Engl. J. Med.* **386**, 509–520 (2022).

14. Li, G., Hilgenfeld, R., Whitley, R. & De Clercq, E. Therapeutic strategies for COVID-19: progress and lessons learned. *Nat. Rev. Drug Discov.* **22**, 449–475 (2023).
15. Lauer, S. A. *et al.* The Incubation Period of Coronavirus Disease 2019 (COVID-19) From Publicly Reported Confirmed Cases: Estimation and Application. *Ann. Intern. Med.* **172**, 577–582 (2020).
16. Dave, N. *et al.* Nosocomial SARS-CoV-2 infections and mortality during unique COVID-19 epidemic waves. *JAMA Netw. Open* **6**, e2341936 (2023).
17. Ponsford, M. J. *et al.* A systematic review and meta-analysis of inpatient mortality associated with nosocomial and community COVID-19 exposes the vulnerability of immunosuppressed adults. *Front. Immunol.* **12**, 744696 (2021).
18. COVID-19: infection prevention and control (IPC). *GOV.UK* https://www.gov.uk/government/publications/wuhan-novel-coronavirus-infection-prevention-and-control (2020).
19. Eyre, D. W. *et al.* Differential occupational risks to healthcare workers from SARS-CoV-2 observed during a prospective observational study. *Elife* **9**, (2020).
20. Voysey, M. *et al.* Safety and efficacy of the ChAdOx1 nCoV-19 vaccine (AZD1222) against SARS-CoV-2: an interim analysis of four randomised controlled trials in Brazil, South Africa, and the UK. *Lancet* **397**, 99–111 (2021).
21. Polack, F. P. *et al.* Safety and Efficacy of the BNT162b2 mRNA Covid-19 Vaccine. *N. Engl. J. Med.* **383**, 2603–2615 (2020).
22. Nguyen, L. H. *et al.* Risk of COVID-19 among front-line health-care workers and the general community: a prospective cohort study. *Lancet Public Health* **5**, e475–e483 (2020).
23. Public Health England. Disparities in the risk and outcomes of COVID August 2020. (2020).
24. SIREN Study Portal. https://snapsurvey.phe.org.uk/siren/.

25. Hall, V. *et al.* Protection against SARS-CoV-2 after covid-19 vaccination and previous infection. *N. Engl. J. Med.* **386**, 1207–1220 (2022).
26. Polaris Observatory Collaborators. Global prevalence, treatment, and prevention of hepatitis B virus infection in 2016: a modelling study. *Lancet Gastroenterol Hepatol* **3**, 383–403 (2018).
27. Global hepatitis report 2024: action for access in low- and middle-income countries. https://www.who.int/publications/i/item/9789240091672 (2024).
28. Devarbhavi, H. *et al.* Global burden of liver disease: 2023 update. *J. Hepatol.* **79**, 516–537 (2023).
29. Global hepatitis report 2024: action for access in low- and middle-income countries. https://www.who.int/publications/i/item/9789240091672 (2024).
30. Global hepatitis report 2024: action for access in low- and middle-income countries. https://www.who.int/publications/i/item/9789240091672 (2024).
31. Hepatitis, V. Global health sector strategy on. https://iris.who.int/bitstream/handle/10665/246177/WHO-HIV-2016.06-eng.pdf?sequence=1.
32. Delphin, M. *et al.* Under-representation of the WHO African region in clinical trials of interventions against hepatitis B virus infection. *Lancet Gastroenterol Hepatol* **9**, 383–392 (2024).
33. O'Hara, G. A. *et al.* Hepatitis B virus infection as a neglected tropical disease. *PLoS Negl. Trop. Dis.* **11**, e0005842 (2017).
34. CDC. Clinical Overview of. *Hepatitis B* https://www.cdc.gov/hepatitis-b/hcp/clinical-overview/index.html#:~:text=Approximately%2015%25%E2%80%9325%25%20of,liver%20failure%2C%20or%20liver%20cancer. (2024).
35. The International Agency for Research on Cancer (IARC). Global Cancer Observatory.

https://gco.iarc.fr/en.

36. Tu, T., Block, J. M., Wang, S., Cohen, C. & Douglas, M. W. The lived experience of chronic hepatitis B: A broader view of its impacts and why we need a cure. *Viruses* **12**, 515 (2020).
37. Ibrahim, Y., Umstead, M., Wang, S. & Cohen, C. The impact of living with chronic hepatitis B on quality of life: Implications for clinical management. *J. Patient Exp.* **10**, 23743735231211069 (2023).
38. Tsukuda, S. & Watashi, K. Hepatitis B virus biology and life cycle. *Antiviral Res.* **182**, 104925 (2020).
39. Kumar, R. Review on hepatitis B virus precore/core promoter mutations and their correlation with genotypes and liver disease severity. *World J. Hepatol.* **14**, 708–718 (2022).
40. Pollicino, T., Cacciola, I., Saffioti, F. & Raimondo, G. Hepatitis B virus PreS/S gene variants: pathobiology and clinical implications. *J. Hepatol.* **61**, 408–417 (2014).
41. Osiowy, C., Giles, E., Tanaka, Y., Mizokami, M. & Minuk, G. Y. Molecular evolution of hepatitis B virus over 25 years. *J. Virol.* **80**, 10307–10314 (2006).
42. Zhou, Y. & Holmes, E. C. Bayesian estimates of the evolutionary rate and age of hepatitis B virus. *J. Mol. Evol.* **65**, 197–205 (2007).
43. Tong, S. & Revill, P. Overview of hepatitis B viral replication and genetic variability. *J. Hepatol.* **64**, S4–S16 (2016).
44. Rajoriya, N., Combet, C., Zoulim, F. & Janssen, H. L. A. How viral genetic variants and genotypes influence disease and treatment outcome of chronic hepatitis B. Time for an individualised approach? *J. Hepatol.* **67**, 1281–1297 (2017).
45. Okamoto, H. *et al.* Typing hepatitis B virus by homology in nucleotide sequence: comparison of surface antigen subtypes. *J. Gen. Virol.* **69 (Pt 10)**, 2575–2583 (1988).
46. Potter, B. I. *et al.* Contemporary and historical human migration patterns shape hepatitis

B virus diversity. *Virus Evol.* **10**, veae009 (2024).

47. Kafeero, H. M. *et al.* Mapping hepatitis B virus genotypes on the African continent from 1997 to 2021: a systematic review with meta-analysis. *Sci. Rep.* **13**, 5723 (2023).
48. Bollyky, P. L., Rambaut, A., Harvey, P. H. & Holmes, E. C. Recombination between sequences of hepatitis B virus from different genotypes. *J. Mol. Evol.* **42**, 97–102 (1996).
49. Simmonds, P. & Midgley, S. Recombination in the genesis and evolution of hepatitis B virus genotypes. *J. Virol.* **79**, 15467–15476 (2005).
50. Pourkarim, M. R., Amini-Bavil-Olyaee, S., Kurbanov, F., Van Ranst, M. & Tacke, F. Molecular identification of hepatitis B virus genotypes/subgenotypes: revised classification hurdles and updated resolutions. *World J. Gastroenterol.* **20**, 7152–7168 (2014).
51. Kramvis, A. & Kew, M. C. Epidemiology of hepatitis B virus in Africa, its genotypes and clinical associations of genotypes. *Hepatol. Res.* **37**, S9–S19 (2007).
52. Kafeero, H. M. *et al.* Disproportionate distribution of HBV genotypes A and D and the recombinant genotype D/E in the high and low HBV endemic regions of Uganda: A wake-up call for regional specific HBV management. *Int. J. Hepatol.* **2022**, 3688547 (2022).
53. McNaughton, A. L., Revill, P. A., Littlejohn, M., Matthews, P. C. & Ansari, M. A. Analysis of genomic-length HBV sequences to determine genotype and subgenotype reference sequences. *J. Gen. Virol.* **101**, 271–283 (2020).
54. Raimondo, G. *et al.* Statements from the Taormina expert meeting on occult hepatitis B virus infection. *J. Hepatol.* **49**, 652–657 (2008).
55. Mak, L.-Y. *et al.* Occult hepatitis B infection and hepatocellular carcinoma: Epidemiology, virology, hepatocarcinogenesis and clinical significance. *J. Hepatol.* **73**, 952–964 (2020).
56. Fu, M. X., Simmonds, P., Andersson, M. & Harvala, H. Biomarkers of transfusion transmitted occult hepatitis B virus infection: Where are we and what next? *Rev. Med.*

Virol. **34**, e2525 (2024).

57. European Association for the Study of the Liver. Electronic address: easloffice@easloffice.eu & European Association for the Study of the Liver. EASL 2017 Clinical Practice Guidelines on the management of hepatitis B virus infection. *J. Hepatol.* **67**, 370–398 (2017).
58. Guidelines for the prevention, diagnosis, care and treatment for people with chronic hepatitis B infection. https://www.who.int/publications/i/item/9789240090903 (2024).
59. Li, Y. *et al.* Systematic review with meta-analysis: the diagnostic accuracy of transient elastography for the staging of liver fibrosis in patients with chronic hepatitis B. *Aliment. Pharmacol. Ther.* **43**, 458–469 (2016).
60. Minier, N. *et al.* Development and evaluation of a simple treatment eligibility score (HEPSANET) to decentralise hepatitis B care in Africa: a cross-sectional study. *Lancet Gastroenterol. Hepatol.* **9**, 323–332 (2024).
61. Wai, C.-T. *et al.* A simple noninvasive index can predict both significant fibrosis and cirrhosis in patients with chronic hepatitis C. *Hepatology* **38**, 518–526 (2003).
62. Hepatitis B. https://www.who.int/news-room/fact-sheets/detail/hepatitis-b.
63. Beasley, R. P., Trepo, C., Stevens, C. E. & Szmuness, W. The e antigen and vertical transmission of hepatitis B surface antigen. *Am. J. Epidemiol.* **105**, 94–98 (1977).
64. Wong, V. C. *et al.* Prevention of the HBsAg carrier state in newborn infants of mothers who are chronic carriers of HBsAg and HBeAg by administration of hepatitis-B vaccine and hepatitis-B immunoglobulin. Double-blind randomised placebo-controlled study. *Lancet* **1**, 921–926 (1984).
65. Okada, K., Kamiyama, I., Inomata, M., Imai, M. & Miyakawa, Y. e antigen and anti-e in the serum of asymptomatic carrier mothers as indicators of positive and negative transmission of hepatitis B virus to their infants. *N. Engl. J. Med.* **294**, 746–749 (1976).
66. Hepatitis: Preventing mother-to-child transmission of the hepatitis B virus.

https://www.who.int/news-room/questions-and-answers/item/hepatitis-preventing-mother-to-child-transmission-of-the-hepatitis-b-virus.

67. Tacke, F. *et al.* Acute hepatitis B virus infection by genotype F despite successful vaccination in an immune-competent German patient. *J. Clin. Virol.* **38**, 353–357 (2007).
68. *Prevention of Mother-to-Child Transmission of Hepatitis B Virus: Guidelines on Antiviral Prophylaxis in Pregnancy*. (World Health Organization, Geneva).
69. Li, X. *et al.* Estimating the health impact of vaccination against ten pathogens in 98 low-income and middle-income countries from 2000 to 2030: a modelling study. *Lancet* **397**, 398–408 (2021).
70. Global Vaccine Alliance to deploy six-in-one vaccine to lower-income countries, establish innovative mechanisms to protect against future epidemic threats. https://www.gavi.org/news/media-room/global-vaccine-alliance-deploy-six-one-vaccine-lower-income-countries (2023).
71. Hepatitis B vaccine. *nhs.uk* https://www.nhs.uk/vaccinations/hepatitis-b-vaccine/.
72. Terrault, N. A. *et al.* Update on prevention, diagnosis, and treatment of chronic hepatitis B: AASLD 2018 hepatitis B guidance. *Hepatology* **67**, 1560–1599 (2018).
73. Hepatitis B EASL guidelines - EASL-the home of hepatology. *EASL-The Home of Hepatology* https://easl.eu/publication/easl-guidelines-management-of-hepatitis-b/ (2017).
74. *Guidelines for the Prevention, Care and Treatment of Persons with Chronic Hepatitis B Infection*. (World Health Organization, Geneva).
75. *Guidelines for the Prevention, Care and Treatment of Persons with Chronic Hepatitis B Infection*. (World Health Organization, Geneva).
76. Lok, A. S. *et al.* Efficacy of entecavir with or without tenofovir disoproxil fumarate for nucleos(t)ide-naïve patients with chronic hepatitis B. *Gastroenterology* **143**, 619–628.e1 (2012).

77. Guidelines for the prevention, diagnosis, care and treatment for people with chronic hepatitis B infection. https://www.who.int/publications/i/item/9789240090903 (2024).
78. Lok, A. S. *et al.* Antiviral drug-resistant HBV: standardization of nomenclature and assays and recommendations for management. *Hepatology* **46**, 254–265 (2007).
79. You, H. *et al.* Guidelines for the Prevention and Treatment of Chronic Hepatitis B (version 2022). *J Clin Transl Hepatol* **11**, 1425–1442 (2023).
80. PROTOCOLO CLÍNICO E DIRETRIZES TERAPÊUTICAS DE HEPATITE B E COINFECÇÕES. Brasília - 2023. https://bvsms.saude.gov.br/bvs/publicacoes/protocolo_clinico_diretrizes_terapeuticas_hepbdigital.pdf.
81. Shimakawa, Y. *et al.* Development of a simple score based on HBeAg and ALT for selecting patients for HBV treatment in Africa. *J. Hepatol.* **69**, 776–784 (2018).
82. Liaw, Y.-F. *et al.* Lamivudine for patients with chronic hepatitis B and advanced liver disease. *N. Engl. J. Med.* **351**, 1521–1531 (2004).
83. Lok, A. S. F. *et al.* Long-term safety of lamivudine treatment in patients with chronic hepatitis B. *Gastroenterology* **125**, 1714–1722 (2003).
84. Nafa, S. *et al.* Early detection of viral resistance by determination of hepatitis B virus polymerase mutations in patients treated by lamivudine for chronic hepatitis B. *Hepatology* **32**, 1078–1088 (2000).
85. Di Marco, V. *et al.* Clinical outcome of HBeAg-negative chronic hepatitis B in relation to virological response to lamivudine. *Hepatology* **40**, 883–891 (2004).
86. Mokaya, J. *et al.* Global prevalence and phylogeny of hepatitis B virus (HBV) drug and vaccine resistance mutations. *J. Viral Hepat.* **28**, 1110–1120 (2021).
87. Mokaya, J. *et al.* Hepatitis B virus resistance to tenofovir: fact or fiction? A systematic literature review and structural analysis of drug resistance mechanisms. *Wellcome Open Res* **5**, 151 (2020).

88. Park, J. W. *et al.* Long-term efficacy of entecavir in adefovir-refractory chronic hepatitis B patients with prior lamivudine resistance. *J. Viral Hepat.* **18**, e475–81 (2011).
89. Hongthanakorn, C. *et al.* Virological breakthrough and resistance in patients with chronic hepatitis B receiving nucleos(t)ide analogues in clinical practice. *Hepatology* **53**, 1854–1863 (2011).
90. Xu, J.-H. *et al.* One hundred and ninety-two weeks treatment of entecavir maleate for Chinese chronic hepatitis B predominantly genotyped B or C. *World J Clin Cases* **10**, 10085–10096 (2022).
91. Tenney, D. J. *et al.* Long-term monitoring shows hepatitis B virus resistance to entecavir in nucleoside-naïve patients is rare through 5 years of therapy. *Hepatology* **49**, 1503–1514 (2009).
92. Jonas, M. M. *et al.* Randomized, controlled trial of entecavir versus placebo in children with hepatitis B envelope antigen-positive chronic hepatitis B. *Hepatology* **63**, 377–387 (2016).
93. Kim, H. J., Cho, Y. K., Jeon, W. K. & Kim, B. I. Clinical characteristics of patients with chronic hepatitis B who developed genotypic resistance to entecavir: Real-life experience. *Clin. Mol. Hepatol.* **23**, 323–330 (2017).
94. Clements, C. J. *et al.* Global control of hepatitis B virus: does treatment-induced antigenic change affect immunization? *Bull. World Health Organ.* **88**, 66–73 (2010).
95. Haché, C. & Villeneuve, J.-P. Lamivudine treatment in patients with chronic hepatitis B and cirrhosis. *Expert Opin. Pharmacother.* **7**, 1835–1843 (2006).
96. Zoulim, F. & Locarnini, S. Hepatitis B virus resistance to nucleos(t)ide analogues. *Gastroenterology* **137**, 1593–608.e1–2 (2009).
97. Guan, R. Efficacy and safety of 5 years lamivudine treatment of Chinese patients with chronic hepatitis B. *J. Gastroenterol. Hepatol.* **16**, 60 (2001).
98. Mokaya, J. *et al.* A blind spot? Confronting the stigma of hepatitis B virus (HBV) infection

- A systematic review. *Wellcome Open Res* **3**, 29 (2018).

99. Liu, K. *et al.* A five years study of antiviral effect of entecavir in Chinese chronic hepatitis B patients. *Sci. Rep.* **6**, 28779 (2016).
100. Lumley, S. F. *et al.* A systematic review and meta-analysis of the risk of hepatitis B virus (HBV) resistance in people treated with entecavir or tenofovir. *J. Clin. Virol.* **174**, 105711 (2024).
101. Lee, J.-H. *et al.* Prior exposure to lamivudine increases entecavir resistance risk in chronic hepatitis B Patients without detectable lamivudine resistance. *Antimicrob. Agents Chemother.* **58**, 1730–1737 (2014).
102. Fung, S. *et al.* Tenofovir disoproxil fumarate (TDF) vs. emtricitabine (FTC)/TDF in lamivudine resistant hepatitis B: A 5-year randomised study. *J. Hepatol.* **66**, 11–18 (2017).
103. Chan, H. L. Y. *et al.* Long-term Treatment With Tenofovir Alafenamide for Chronic Hepatitis B Results in High Rates of Viral Suppression and Favorable Renal and Bone Safety. *Am. J. Gastroenterol.* (2023) doi:10.14309/ajg.0000000000002468.
104. Mokaya, J. *et al.* A systematic review of hepatitis B virus (HBV) drug and vaccine escape mutations in Africa: A call for urgent action. *PLoS Negl. Trop. Dis.* **12**, e0006629 (2018).
105. Guirgis, B. S. S., Abbas, R. O. & Azzazy, H. M. E. Hepatitis B virus genotyping: current methods and clinical implications. *Int. J. Infect. Dis.* **14**, e941–53 (2010).
106. Garcia-Garcia, S. *et al.* Next-generation sequencing for the diagnosis of hepatitis B: current status and future prospects. *Expert Rev. Mol. Diagn.* **21**, 381–396 (2021).
107. Houldcroft, C. J., Beale, M. A. & Breuer, J. Clinical and biological insights from viral genome sequencing. *Nat. Rev. Microbiol.* **15**, 183–192 (2017).
108. Downs, L. O. *et al.* Bimodal distribution and set point HBV DNA viral loads in chronic infection: retrospective analysis of cohorts from the UK and South Africa. *Wellcome Open Res* **5**, 113 (2020).

109. Artic Network. https://artic.network/ncov-2019.

110. Quick, J. *et al.* Real-time, portable genome sequencing for Ebola surveillance. *Nature* **530**, 228–232 (2016).

111. Downs, L. O., Eyre, D. W., O'Donnell, D. & Jeffery, K. Home-based SARS-CoV-2 lateral flow antigen testing in hospital workers. *J. Infect.* **82**, 282–327 (2021).

112. *nCoV-2019/V3 at Master · Artic-Network/primer-Schemes*. (Github).

113. Freed, N. E., Vlková, M., Faisal, M. B. & Silander, O. K. Rapid and inexpensive whole-genome sequencing of SARS-CoV-2 using 1200 bp tiled amplicons and Oxford Nanopore Rapid Barcoding. *Biol. Methods Protoc.* **5**, bpaa014 (2020).

114. Quick, J. nCoV-2019 sequencing protocol v3 (LoCost). (2020).

115. Loman, N. *et al. Artic-Network Field Bioinformatics: 1.2.1*. (2021). doi:10.5281/zenodo.4441073.

116. Moshiri, N. *TreeSwift*. (Github).

117. Eaton, D. A. R. Toytree: A minimalist tree visualization and manipulation library for Python. *Methods Ecol. Evol.* **11**, 187–191 (2020).

118. Ripley, B. D. The R project in statistical computing. *MSOR connect.* **1**, 23–25 (2001).

119. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis*. (Springer, Cham, 2016).

120. igraph – Network analysis software. https://igraph.org.

121. Bannister, E. *et al.* Analysis of the in vitro replication phenotype of African hepatitis B virus (HBV) genotypes and subgenotypes present in Australia identifies marked differences in DNA and protein expression. *Virology* **540**, 97–103 (2020).

122. Downs, L. O. *et al.* Electronic Health Informatics Data To Describe Clearance Dynamics of Hepatitis B Surface Antigen (HBsAg) and e Antigen (HBeAg) in Chronic Hepatitis B Virus Infection. *MBio* **10**, (2019).

123. O'Hara, G. *et al.* Liver function tests and fibrosis scores in a rural population in Africa: a cross-sectional study to estimate the burden of disease and associated risk factors. *BMJ*

Open **10**, e032890 (2020).

124. Asiki, G. *et al.* The general population cohort in rural south-western Uganda: a platform for communicable and non-communicable disease studies. *Int. J. Epidemiol.* **42**, 129–141 (2013).

125. Morgan, C. E. *et al.* Hepatitis B virus prevalence and transmission in the households of pregnant women in Kinshasa, Democratic Republic of Congo. *Open Forum Infect. Dis.* **11**, ofae150 (2024).

126. Downs, L. O. *et al.* STRIKE-HBV: establishing an HBV screening programme in Kilifi, Kenya-challenges, successes and lessons learnt. *Sex. Transm. Infect.* **100**, 325–328 (2024).

127. Fu, M. X. *et al.* Ultrasensitive PCR system for HBV DNA detection: Risk stratification for occult hepatitis B virus infection in English blood donors. *J. Med. Virol.* **95**, e29144 (2023).

128. https://theses.hal.science/tel-03678598v1.

129. Simon, N., Shallat, J., Williams Wietzikoski, C. & Harrington, W. E. Optimization of Chelex 100 resin-based extraction of genomic DNA from dried blood spots. *Biol. Methods Protoc.* **5**, bpaa009 (2020).

130. Margeridon, S. *et al.* Rolling circle amplification, a powerful tool for genetic and functional studies of complete hepatitis B virus genomes from low-level infections and for directly probing covalently closed circular DNA. *Antimicrob. Agents Chemother.* **52**, 3068–3073 (2008).

131. Günther, S. *et al.* A novel method for efficient amplification of whole hepatitis B virus genomes permits rapid functional analysis and reveals deletion mutants in immunosuppressed patients. *J. Virol.* **69**, 5437–5444 (1995).

132. Ansari, A. HBV probe sequences. figshare https://doi.org/10.6084/M9.FIGSHARE.22127015.V1 (2023).

133. Lumley, S. F. *et al.* Pan-genotypic probe-based enrichment to improve efficiency of Hepatitis B virus sequencing. *bioRxiv* 2023.02.20.529276 (2023) doi:10.1101/2023.02.20.529276.

134. Bonsall, D. *et al.* ve-SEQ: Robust, unbiased enrichment for streamlined detection and whole-genome sequencing of HCV and other highly diverse pathogens. *F1000Res.* **4**, 1062 (2015).

135. Quick, J. *et al.* Multiplex PCR method for MinION and Illumina sequencing of Zika and other virus genomes directly from clinical samples. *Nat. Protoc.* **12**, 1261–1276 (2017).

136. Hayer, J. *et al.* HBVdb: a knowledge database for Hepatitis B Virus. *Nucleic Acids Res.* **41**, D566–70 (2013).

137. Katoh, K. & Toh, H. Parallelization of the MAFFT multiple sequence alignment program. *Bioinformatics* **26**, 1899–1900 (2010).

138. Menardo, F. *et al.* Treemmer: a tool to reduce large phylogenetic datasets with minimal loss of diversity. *BMC Bioinformatics* **19**, 164 (2018).

139. *primerschemes/hbv/500/v1.1.0 at Main · Quick-Lab/primerschemes*. (Github).

140. Chen, Y., Sze, J. & He, M.-L. HBV cccDNA in patients' sera as an indicator for HBV reactivation and an early signal of liver damage. *World J. Gastroenterol.* **10**, 82–85 (2004).

141. Takkenberg, R. B. *et al.* Validation of a sensitive and specific real-time PCR for detection and quantitation of hepatitis B virus covalently closed circular DNA in plasma of chronic hepatitis B patients. *J. Med. Virol.* **81**, 988–995 (2009).

142. *primerschemes/hbv/600/v2.1.0 at Main · Quick-Lab/primerschemes*. (Github).

143. McNaughton, A. L. *et al.* Illumina and Nanopore methods for whole genome sequencing of hepatitis B virus (HBV). *Sci. Rep.* **9**, 7081 (2019).

144. Martel, N., Gomes, S. A., Chemin, I., Trépo, C. & Kay, A. Improved rolling circle amplification (RCA) of hepatitis B virus (HBV) relaxed-circular serum DNA (RC-DNA). *J.*

Virol. Methods **193**, 653–659 (2013).

145. Limberis, J. D. uDumBell – Circularization of rv0678 for genotypic bedaquiline resistance testing of Mycobacterium tuber. (2023).

146. Lumley, S. F. *et al.* Whole genome sequencing of hepatitis B virus (HBV) using tiled amplicon (HEP-TILE) and probe-based enrichment on Illumina and Nanopore platforms. *medRxiv* 2024.09.11.24313306 (2024) doi:10.1101/2024.09.11.24313306.

147. New England Biolabs. NEBNext® Ultra™ II FS DNA Library Prep Kit for Illumina. https://www.neb.com/en-gb/products/e7805-nebnext-ultra-ii-fs-dna-library-prep-kit-for-illumina?srsltid=AfmBOoqLScj76UVKUy3_k2XThERSPIrso5ZR-LhfqKfMZN4tQ6nQ5gkj.

148. Quick, J. nCoV-2019 sequencing protocol v3 (LoCost) v3. (2020) doi:10.17504/protocols.io.bp2l6n26rgqe/v3.

149. Lumley, S., Kent, C., Quick, J. & Matthews, P. HEP-TILE: HBV whole genome sequencing (nanopore protocol) v1. (2024) doi:10.17504/protocols.io.5jyl82bedl2w/v1.

150. Creators nickloman Will Rowe1 Radoslaw Poplawski Stephen Rudd2 Sam W3 ChrisKBio3 Andrew Rambaut Trevor Bedford4 Karthik5 Jared Simpson Josh Quick3 mattloose Show affiliations 1. @grailbio 2. @nanoporetech 3. University of Birmingham 4. Fred Hutchinson Cancer Research Center 5. The Scripps Research Institute. *ChrisgKent/hbv-Fieldbioinfomatics: Zenodo Release*. doi:10.5281/zenodo.12626842.

151. MEGA. https://www.megasoftware.net/.

152. Geneious. *Geneious* https://www.geneious.com/.

153. Llc, G. B. SnapGene. *SnapGene* https://www.snapgene.com/.

154. IGV: Integrative Genomics Viewer. https://igv.org/.

155. IQTREE Web Server: Fast and accurate phylogenetic trees under maximum likelihood. http://iqtree.cibiv.univie.ac.at/.

156. FigTree. http://tree.bio.ed.ac.uk/software/figtree/.

157. McNaughton, A., Ansari, A. & Matthews, P. Hepatitis B Virus (HBV) Genotype and

Subtype Reference Sequences. figshare https://doi.org/10.6084/M9.FIGSHARE.8851946.V1 (2019).

158. Genotyping. https://www.ncbi.nlm.nih.gov/projects/genotyping/formpage.cgi.

159. Geno2pheno hbv. https://hbv.geno2pheno.org/index.php.

160. Biostrings. *Bioconductor* https://bioconductor.org/packages/release/bioc/html/Biostrings.html.

161. Tidyverse. https://www.tidyverse.org/.

162. Crooks, G. E. WebLogo - Create Sequence Logos. https://weblogo.berkeley.edu/logo.cgi.

163. Park, E.-S. *et al.* Identification of a quadruple mutation that confers tenofovir resistance in chronic hepatitis B patients. *J. Hepatol.* **70**, 1093–1102 (2019).

164. Sheldon, J. *et al.* Selection of hepatitis B virus polymerase mutations in HIV-coinfected patients treated with tenofovir. *Antivir. Ther.* **10**, 727–734 (2005).

165. Lada, O. *et al.* Quasispecies analysis and in vitro susceptibility of HBV strains isolated from HIV-HBV-coinfected patients with delayed response to tenofovir. *Antivir. Ther.* **17**, 61–70 (2012).

166. Lumley, S. F. *et al.* Epidemiological data and genome sequencing reveals that nosocomial transmission of SARS-CoV-2 is underestimated and mostly mediated by a small number of highly infectious individuals. *J. Infect.* **83**, 473–482 (2021).

167. Epidemiological data and genome sequencing reveals that nosocomial transmission of SARS-CoV-2 is underestimated and mostly mediated by a small number of highly infectious individuals. https://www.altmetric.com/details/110937536/policy-documents.

168. Lumley, S. Multiple sequence alignment. figshare https://doi.org/10.6084/M9.FIGSHARE.14883213.V2 (2021).

169. Lumley, S. Maximum likelihood phylogeny. figshare https://doi.org/10.6084/M9.FIGSHARE.14883216.V1 (2021).

170. SARS-CoV-2 in Oxfordshire hospitals ’20-21. *Microreact* https://microreact.org/project/sroEhcsq8kkfNXTorBZaNb?&tal=1&tns=10&tts=8&mns=100&uvs=0&uhs=70&lns=4&tc=1&tl=1&lu=w&fb=lineage__autocolour,nosocomial_acquisition__autocolour&fl=nosocomial_acquisition__autocolour.

171. https://oxfordbrc.nihr.ac.uk/research-themes-overview/antimicrobial-resistance-and-modernising-microbiology/infections-in-oxfordshire-research-database-iord/.

172. Wang, D. *et al.* Clinical Characteristics of 138 Hospitalized Patients With 2019 Novel Coronavirus-Infected Pneumonia in Wuhan, China. *JAMA* **323**, 1061–1069 (2020).

173. Carter, B. *et al.* Nosocomial COVID-19 infection: examining the risk of mortality. The COPE-Nosocomial Study (COVID in Older PEople). *J. Hosp. Infect.* **106**, 376–384 (2020).

174. Discombe, M. Covid infections caught in hospital rise by a third in one week. https://www.hsj.co.uk/patient-safety/covid-infections-caught-in-hospital-rise-by-a-third-in-one-week/7029211.article.

175. Campbell, D. & Bawden, A. Up to 8,700 patients died after catching Covid in English hospitals. *The Guardian* (2021).

176. Shah, A. S. V. *et al.* Risk of hospital admission with coronavirus disease 2019 in healthcare workers and their households: nationwide linkage cohort study. *BMJ* **371**, m3582 (2020).

177. Stirrup, O. T. *et al.* Rapid feedback on hospital onset SARS-CoV-2 infections combining epidemiological and sequencing data. *bioRxiv* (2020) doi:10.1101/2020.11.12.20230326.

178. Snell, L. B. *et al.* Combined epidemiological and genomic analysis of nosocomial SARS-CoV-2 transmission identifies community social distancing as the dominant intervention reducing outbreaks. *bioRxiv* (2020) doi:10.1101/2020.11.17.20232827.

179. Illingworth, C. Superspreaders drive the largest outbreaks of hospital onset COVID-19

infection. https://osf.io/wmkn3/.

180. Ellingford, J. M. *et al.* Genomic and healthcare dynamics of nosocomial SARS-CoV-2 transmission. *Elife* **10**, (2021).

181. Meredith, L. W. *et al.* Rapid implementation of SARS-CoV-2 sequencing to investigate cases of health-care associated COVID-19: a prospective genomic surveillance study. *Lancet Infect. Dis.* **20**, 1263–1271 (2020).

182. Lucey, M. *et al.* Whole-genome sequencing to track SARS-CoV-2 transmission in nosocomial outbreaks. *Clin. Infect. Dis.* (2020) doi:10.1093/cid/ciaa1433.

183. Hill, V. *et al.* The origins and molecular evolution of SARS-CoV-2 lineage B.1.1.7 in the UK. *Virus Evol.* **8**, veac080 (2022).

184. Tao, K. *et al.* The biological and clinical significance of emerging SARS-CoV-2 variants. *Nat. Rev. Genet.* **22**, 757–773 (2021).

185. Public Health England. COVID-19: infection prevention and control (IPC). https://www.gov.uk/government/publications/wuhan-novel-coronavirus-infection-prevention-and-control (2020).

186. *nCoV-2019/V3 at Master · Artic-Network/primer-Schemes*. (Github).

187. Freed, N. E., Vlková, M., Faisal, M. B. & Silander, O. K. Rapid and inexpensive whole-genome sequencing of SARS-CoV-2 using 1200 bp tiled amplicons and Oxford Nanopore rapid barcoding. *bioRxiv* (2020) doi:10.1101/2020.05.28.122648.

188. Wilcox, M. Correct measurement and reporting of healthcare associated-COVID-19 cases: letter to Regional Chief Nurses, Regional Medical Directors. (2020).

189. Ferretti, L. *et al.* Quantifying SARS-CoV-2 transmission suggests epidemic control with digital contact tracing. *Science* **368**, (2020).

190. Griffin, J. *et al.* Rapid review of available evidence on the serial interval and generation time of COVID-19. *BMJ Open* **10**, e040263 (2020).

191. Park, M., Cook, A. R., Lim, J. T., Sun, Y. & Dickens, B. L. A Systematic Review of

COVID-19 Epidemiology Based on Current Evidence. *J. Clin. Med. Res.* **9**, (2020).

192. Duchene, S. *et al.* Temporal signal and the phylodynamic threshold of SARS-CoV-2. *Virus Evol* **6**, veaa061 (2020).

193. COVID-19: epidemiological definitions of outbreaks and clusters in particular settings. https://www.gov.uk/government/publications/covid-19-epidemiological-definitions-of-outbreaks-and-clusters/covid-19-epidemiological-definitions-of-outbreaks-and-clusters-in-particular-settings.

194. Eyre, D. W. *et al.* Diverse sources of C. difficile infection identified on whole-genome sequencing. *N. Engl. J. Med.* **369**, 1195–1205 (2013).

195. Price, J. R. *et al.* Transmission of Staphylococcus aureus between health-care workers, the environment, and patients in an intensive care unit: a longitudinal cohort study based on whole-genome sequencing. *Lancet Infect. Dis.* **17**, 207–214 (2017).

196. Lemieux, J. E. *et al.* Phylogenetic analysis of SARS-CoV-2 in Boston highlights the impact of superspreading events. *Science* **371**, (2021).

197. Adam, D. C. *et al.* Clustering and superspreading potential of SARS-CoV-2 infections in Hong Kong. *Nat. Med.* **26**, 1714–1719 (2020).

198. Endo, A., Centre for the Mathematical Modelling of Infectious Diseases COVID-19 Working Group, Abbott, S., Kucharski, A. J. & Funk, S. Estimating the overdispersion in COVID-19 transmission using outbreak sizes outside China. *Wellcome Open Res* **5**, 67 (2020).

199. Popa, A. *et al.* Genomic epidemiology of superspreading events in Austria reveals mutational dynamics and transmission properties of SARS-CoV-2. *Sci. Transl. Med.* **12**, (2020).

200. Hamilton, W. L. *et al.* Applying prospective genomic surveillance to support investigation of hospital-onset COVID-19. *Lancet Infect. Dis.* (2021) doi:10.1016/S1473-3099(21)00251-6.

201. San, J. E. *et al.* Transmission dynamics of SARS-CoV-2 within-host diversity in two major hospital outbreaks in South Africa. *Virus Evol* **7**, veab041 (2021).
202. Borges, V. *et al.* Nosocomial Outbreak of SARS-CoV-2 in a 'Non-COVID-19' Hospital Ward: Virus Genome Sequencing as a Key Tool to Understand Cryptic Transmission. *Viruses* **13**, (2021).
203. Kissler, S. M. *et al.* SARS-CoV-2 viral dynamics in acute infections. *bioRxiv* (2020) doi:10.1101/2020.10.21.20217042.
204. Mo, Y. *et al.* Transmission dynamics of SARS-CoV-2 in the hospital setting. *bioRxiv* (2021) doi:10.1101/2021.04.28.21256245.
205. Blackstone, J. *et al.* Protocol for the COG-UK hospital-onset COVID-19 infection (HOCI) multicentre interventional clinical study: evaluating the efficacy of rapid genome sequencing of SARS-CoV-2 in limiting the spread of COVID-19 in UK NHS hospitals. *BMJ Open* **12**, e052514 (2022).
206. Panca, M. *et al.* Evaluating the cost implications of integrating SARS-CoV-2 genome sequencing for infection prevention and control investigation of nosocomial transmission within hospitals. *J. Hosp. Infect.* **139**, 23–32 (2023).
207. Stirrup, O. *et al.* Effectiveness of rapid SARS-CoV-2 genome sequencing in supporting infection control for hospital-onset COVID-19 infection: Multicentre, prospective study. *Elife* **11**, e78427 (2022).
208. Lumley, S. F. *et al.* Whole genome sequencing of hepatitis B virus (HBV) using tiled amplicon (HEP-TILE) and probe-based enrichment on Illumina and Nanopore platforms. *Infectious Diseases (except HIV/AIDS)* (2024).
209. Lumley, S. *et al.* 2023 iHBV meeting poster: Full length HBV genomes in a single read using Oxford Nanopore Technologies MinION. (2024) doi:10.6084/M9.FIGSHARE.25379503.V1.
210. Lumley, S., Matthews, P. & Delphin, M. 2023 iHBV poster: Pan-genotypic probe-based

enrichment to improve efficiency of Hepatitis B virus sequencing. (2024) doi:10.6084/M9.FIGSHARE.25379560.V1.

211. Lumley, S. & Matthews, P. 2024 EASL HEP-TILE HBV sequencing poster. (2024) doi:10.6084/M9.FIGSHARE.25909276.V1.

212. Lok, A. S. Personalized treatment of hepatitis B. *Clin. Mol. Hepatol.* **21**, 1–6 (2015).

213. Wang, X. *et al.* Integrating nested PCR with high-throughput sequencing to characterize mutations of HBV genome in low viral load samples. *Medicine* **96**, e7588 (2017).

214. Rector, A., Tachezy, R. & Van Ranst, M. A sequence-independent strategy for detection and cloning of circular DNA virus genomes by using multiply primed rolling-circle amplification. *J. Virol.* **78**, 4993–4998 (2004).

215. Singla, B. *et al.* Levels of hepatitis B virus replicative intermediate in serum samples of chronic hepatitis B patients. *Mol. Biol. Rep.* **41**, 4689–4696 (2014).

216. Wei, L. & Ploss, A. Hepatitis B virus cccDNA is formed through distinct repair processes of each strand. *Nat. Commun.* **12**, 1591 (2021).

217. Wei, L. & Ploss, A. Mechanism of hepatitis B virus cccDNA formation. *Viruses* **13**, 1463 (2021).

218. Briese, T. *et al.* Virome Capture Sequencing Enables Sensitive Viral Diagnosis and Comprehensive Virome Analysis. *MBio* **6**, e01491–15 (2015).

219. Depledge, D. P. *et al.* Specific Capture and Whole-Genome Sequencing of Viruses from Clinical Samples. *PLoS ONE* vol. 6 e27805 Preprint at https://doi.org/10.1371/journal.pone.0027805 (2011).

220. Berg, M. G. *et al.* Advanced molecular surveillance approaches for characterization of blood borne hepatitis viruses. *PLoS One* **15**, e0236046 (2020).

221. *Fieldbioinformatics: The ARTIC Field Bioinformatics Pipeline*. (Github).

222. Stenbäck, J. B. *et al.* Accurate and cost-efficient whole genome sequencing of hepatitis B virus using Nanopore. *medRxiv* 2024.08.12.24311345 (2024)

doi:10.1101/2024.08.12.24311345.

223. Ringlander, J., Andersson, M. E., Prakash, K., Larsson, S. B. & Lindh, M. Deep sequencing of hepatitis B virus using Ion Torrent fusion primer method. *J. Virol. Methods* **299**, 114315 (2022).

224. Bezerra, C. S. *et al.* Dried blood spot sampling for hepatitis B virus quantification, sequencing and mutation detection. *Sci. Rep.* **12**, 1651 (2022).

225. Thompson, P. *et al.* Seroepidemiology of hepatitis B in the Democratic Republic of the Congo. *Am. J. Trop. Med. Hyg.* **101**, 226–229 (2019).

226. Stene-Johansen, K. *et al.* Dry Blood Spots a reliable method for measurement of hepatitis B viral load in resource-limited settings. *PLoS One* **11**, e0166201 (2016).

227. Lange, B. *et al.* Diagnostic accuracy of serological diagnosis of hepatitis C and B using dried blood spot samples (DBS): two systematic reviews and meta-analyses. *BMC Infect. Dis.* **17**, 700 (2017).

228. Kikuchi, M. *et al.* Dried blood spot is the feasible matrix for detection of some but not all hepatitis B virus markers of infection. *BMC Res. Notes* **15**, 287 (2022).

229. Astbury, S. *et al.* Nanopore sequencing from extraction-free direct PCR of dried serum spots for portable hepatitis B virus drug-resistance typing. *J. Clin. Virol.* **129**, 104483 (2020).

230. Xu, Y. *et al.* Detection of Viral Pathogens With Multiplex Nanopore MinION Sequencing: Be Careful With Cross-Talk. *Front. Microbiol.* **9**, 2225 (2018).

231. Tshiabuila, D. *et al.* An Oxford Nanopore Technology-based Hepatitis B virus sequencing protocol suitable for genomic surveillance within clinical diagnostic settings. *Int. J. Mol. Sci.* **25**, (2024).

232. Tshiabuila, D. *et al.* An Oxford Nanopore Technology-Based Hepatitis B Virus Sequencing Protocol Suitable For Genomic Surveillance Within Clinical Diagnostic Settings. *medRxiv* (2024) doi:10.1101/2024.01.19.24301519.

233. PrimalScheme: primer panels for multiplex PCR. https://primalscheme.com/.

234. Lumley, S. *et al.* Multiple risk factors for persistent HBV viraemia in an adult receiving nucleos/tide analogue therapy. *Sex. Transm. Infect.* **100**, 329–331 (2024).

235. Wang, T. *et al.* Classification of virologic trajectories during nucleos/tide analogue treatment of hepatitis B virus (HBV) infection. *bioRxiv* (2023) doi:10.1101/2023.12.01.23299288.

236. Hepatitis B (chronic): diagnosis and management | Guidance | NICE (2017).

237. Lumley, S. F. *et al.* A systematic review and meta-analysis of the risk of hepatitis B virus (HBV) genotypic resistance in people treated with entecavir or tenofovir. *bioRxiv* (2023) doi:10.1101/2023.11.08.23298154.

238. Allard, N. L. *et al.* Adherence in chronic hepatitis B: associations between medication possession ratio and adverse viral outcomes. *BMC Gastroenterol.* **20**, 140 (2020).

239. Hunter, J. P. *et al.* In meta-analyses of proportion studies, funnel plots were found to be an inaccurate method of assessing publication bias. *J. Clin. Epidemiol.* **67**, 897–903 (2014).

240. Stuyver, L. J. *et al.* Nomenclature for antiviral-resistant human hepatitis B virus mutations in the polymerase region. *Hepatology* **33**, 751–757 (2001).

241. Hou, J. L. *et al.* Tenofovir disoproxil fumarate vs adefovir dipivoxil in Chinese patients with chronic hepatitis B after 48 weeks: a randomized controlled trial. *J. Viral Hepat.* **22**, 85–93 (2015).

242. Patterson, S. J. *et al.* Tenofovir disoproxil fumarate rescue therapy following failure of both lamivudine and adefovir dipivoxil in chronic hepatitis B. *Gut* **60**, 247–254 (2011).

243. Tan, J. *et al.* Tenofovir monotherapy is effective in hepatitis B patients with antiviral treatment failure to adefovir in the absence of adefovir-resistant mutations. *J. Hepatol.* **48**, 391–398 (2008).

244. De Francesco, M. A. *et al.* Clinical course of chronic hepatitis B patients receiving

nucleos(t)ide analogues after virological breakthrough during monotherapy with lamivudine. *New Microbiol.* **38**, 29–37 (2015).

245. Srivastava, M., Singh, N., Dixit, V. K., Nath, G. & Jain, A. K. Comparative evaluation of long-term monotherapies & combination therapies in patients with chronic hepatitis B: A pilot study. *Indian J. Med. Res.* **144**, 424–432 (2016).

246. Heo, J. *et al.* A 96-week randomized trial of switching to entecavir in chronic hepatitis B patients with a partial virological response to lamivudine. *Antivir. Ther.* **17**, 1563–1570 (2012).

247. Cho, E. J. *et al.* Comparison of the efficacy of entecavir and tenofovir in nucleos(T)ide analogue-experienced chronic hepatitis B patients. *PLoS One* **10**, e0130392 (2015).

248. Wang, J. C., He, L. L. & Chen, Q. Comparison of re-treatment outcomes of lamivudine plus adefovir or entecavir in chronic hepatitis B patients with viral relapse after cessation of adefovir. *Eur. Rev. Med. Pharmacol. Sci.* **17**, 1162–1166 (2013).

249. Suzuki, F. *et al.* Efficacy of switching to entecavir monotherapy in Japanese lamivudine-pretreated patients. *J. Gastroenterol. Hepatol.* **25**, 892–898 (2010).

250. Chang, T.-T. *et al.* A comparison of entecavir and lamivudine for HBeAg-positive chronic hepatitis B. *N. Engl. J. Med.* **354**, 1001–1010 (2006).

251. Guo, X. *et al.* Trends in hepatitis B virus resistance to nucleoside/nucleotide analogues in North China from 2009-2016: A retrospective study. *Int. J. Antimicrob. Agents* **52**, 201–209 (2018).

252. Lim, L. *et al.* Five-year efficacy and safety of tenofovir-based salvage therapy for patients with chronic hepatitis B who previously failed LAM/ADV therapy. *Liver Int.* **37**, 827–835 (2017).

253. Marcellin, P. *et al.* Ten-year efficacy and safety of tenofovir disoproxil fumarate treatment for chronic hepatitis B virus infection. *Liver Int.* **39**, 1868–1875 (2019).

254. Pan, C. Q. *et al.* Similar efficacy and safety of tenofovir in Asians and non-Asians with

chronic hepatitis B. *World J. Gastroenterol.* **21**, 5524–5531 (2015).

255. Sayan, M. & Akhan, S. C. Antiviral drug-associated potential vaccine-escape hepatitis B virus mutants in Turkish patients with chronic hepatitis B. *Int. J. Infect. Dis.* **15**, e722–6 (2011).

256. Sayan, M., Cavdar, C. & Dogan, C. Naturally occurring polymerase and surface gene variants of hepatitis B virus in Turkish hemodialysis patients with chronic hepatitis B. *Jpn. J. Infect. Dis.* **65**, 495–501 (2012).

257. Yamashige, D. *et al.* Effectiveness of tenofovir alafenamide for chronic hepatitis B patients with a poor response to the previously used nucleos(t)ide analogs. *J. Gastroenterol.* **56**, 1008–1021 (2021).

258. Overview | Hepatitis B (chronic): diagnosis and management | Guidance | NICE.

259. Lumley, S. *et al.* Liver elastography scores in a large population cohort in rural Uganda: Uganda Liver Disease Study (ULiDS). (2024) doi:10.6084/M9.FIGSHARE.25377049.V1.

260. Morgan, C. *et al.* WED-469 Investigating household hepatitis B transmission using whole-genome sequencing and serological analysis in Kinshasa Province, Democratic Republic of Congo. *J. Hepatol.* **80**, S675–S676 (2024).

261. Morgan, C. Household hepatitis B virus transmission using whole genome sequencing and serological analysis in Kinshasa Province, Democratic Republic of Congo. https://figshare.com/articles/poster/EASL_2024_poster_Investigating_household_hepatitis_B_virus_transmission_using_whole-genome_sequencing_and_serological_analysis_in_Kinshasa_Province_Democratic_Republic_of_Congo/26065075/1?file=47139661.

262. Guidelines for the prevention, diagnosis, care and treatment for people with chronic hepatitis B infection. https://www.who.int/publications/i/item/9789240090903 (2024).

263. Africa. *MapChart* https://www.mapchart.net/africa.html.

264. Spearman, C. W. *et al.* A new approach to prevent, diagnose, and treat hepatitis B in Africa. *BMC Glob Public Health* **1**, 24 (2023).

265. Lemoine, M. *et al.* The gamma-glutamyl transpeptidase to platelet ratio (GPR) predicts significant liver fibrosis and cirrhosis in patients with chronic HBV infection in West Africa. *Gut* **65**, 1369–1376 (2016).

266. McNaughton, A. L., Lemoine, M., van Rensburg, C. & Matthews, P. C. Extending treatment eligibility for chronic hepatitis B virus infection. *Nat. Rev. Gastroenterol. Hepatol.* **18**, 146–147 (2021).

267. Anderson, M. *et al.* ELASTOGRAPHY AND CAP ASSESSMENT FOR LIVER DISEASE: Training events and feedback representing clinical research programmes in South Africa, Uganda and Kenya. Preprint at https://doi.org/10.6084/M9.FIGSHARE.27941958.V1 (2024).

268. European Association for the Study of the Liver (EASL). Electronic address: easloffice@easloffice.eu, European Association for the Study of Diabetes (EASD), European Association for the Study of Obesity (EASO) & European Association for the Study of the Liver (EASL). EASL-EASD-EASO Clinical Practice Guidelines on the management of metabolic dysfunction-associated steatotic liver disease (MASLD). *J. Hepatol.* **81**, 492–542 (2024).

269. Guidelines for the prevention, diagnosis, care and treatment for people with chronic hepatitis B infection. https://www.who.int/publications/i/item/9789240090903 (2024).

270. Nguyen, L. B. L. *et al.* Treat All versus targeted strategies to select HBV-infected people for antiviral therapy in The Gambia, west Africa: a cost-effectiveness analysis. *Lancet Glob. Health* **12**, e66–e78 (2024).

271. Wang, T. *et al.* Distinct virologic trajectories in chronic hepatitis B identify heterogeneity in response to nucleos/tide analogue therapy. *JHEP Rep.* **0**, 101229 (2024).

272. Marcellin, P. *et al.* Tenofovir disoproxil fumarate versus adefovir dipivoxil for chronic hepatitis B. *N. Engl. J. Med.* **359**, 2442–2455 (2008).

273. Kitrinos, K. M. *et al.* No detectable resistance to tenofovir disoproxil fumarate after 6

years of therapy in patients with chronic hepatitis B. *Hepatology* **59**, 434–442 (2014).

274. Lumley, S., Quick, J., Matthews, P. & Kent, C. HEP-TILE: HBV whole genome sequencing (nanopore protocol). (2024).

275. *primerschemes/hbv/600/v2.1.0 at Main · Quick-Lab/primerschemes*. (Github).

276. Mikulska, M. *et al.* Case report: management and HBV sequencing in a patient co-infected with HBV and HIV failing tenofovir. *J. Med. Virol.* **84**, 1340–1343 (2012).

277. Ghany, M. G. *et al.* Hepatitis B virus S mutants in liver transplant recipients who were reinfected despite hepatitis B immune globulin prophylaxis. *Hepatology* **27**, 213–222 (1998).

278. Cooreman, M. P., Leroux-Roels, G. & Paulij, W. P. Vaccine- and hepatitis B immune globulin-induced escape mutations of hepatitis B virus surface antigen. *J. Biomed. Sci.* **8**, 237–247 (2001).

279. Avellón, A. & Echevarria, J. M. Frequency of hepatitis B virus ‘a’ determinant variants in unselected Spanish chronic carriers. *J. Med. Virol.* **78**, 24–36 (2006).

280. Maponga, T. G. *et al.* A putative hepatitis B virus sequence motif associated with hepatocellular carcinoma in South African adults. *Infectious Diseases (except HIV/AIDS)* (2024).

281. Sali, S., Azarmmanesh, S., Ghalikhani, H. & Vaezjalali, M. Phylogenetic analysis of Hepatitis B virus among household members with HBV chronic infection. *Avicenna J. Med. Biotechnol.* **11**, 221–228 (2019).

282. Zampino, R. *et al.* Intra-familial transmission of hepatitis B virus in Italy: phylogenetic sequence analysis and amino-acid variation of the core gene. *J. Hepatol.* **36**, 248–253 (2002).

283. Datta, S., Banerjee, A., Chandra, P. K., Chowdhury, A. & Chakravarty, R. Genotype, phylogenetic analysis, and transmission pattern of occult hepatitis B virus (HBV) infection in families of asymptomatic HBsAg carriers. *J. Med. Virol.* **78**, 53–59 (2006).

284. Dumpis, U. *et al.* Transmission of hepatitis B virus infection in Gambian families revealed by phylogenetic analysis. *J. Hepatol.* **35**, 99–104 (2001).

285. Sede, M. *et al.* Hepatitis B virus depicts a high degree of conservation during the immune-tolerant phase in familiarly transmitted chronic hepatitis B infection: deep-sequencing and phylogenetic analysis. *J. Viral Hepat.* **21**, 650–661 (2014).

286. Stenbäck, J. B. *et al.* Accurate and cost-efficient whole genome sequencing of hepatitis B virus using Nanopore. *Infectious Diseases (except HIV/AIDS)* (2024).

287. Supasa, P. *et al.* Reduced neutralization of SARS-CoV-2 B.1.1.7 variant by convalescent and vaccine sera. *Cell* **184**, 2201–2211.e7 (2021).

288. Zhou, D. *et al.* Evidence of escape of SARS-CoV-2 variant B.1.351 from natural and vaccine-induced sera. *Cell* (2021) doi:10.1016/j.cell.2021.02.037.

289. Chowdhury, O. *et al.* Impaired antibody response to COVID-19 vaccination in patients with chronic myeloid neoplasms. *Br. J. Haematol.* **194**, 1010–1015 (2021).

290. Shankar, S. *et al.* SARS-CoV-2-specific T cell responses are not associated with protection against reinfection in hemodialysis patients. *J. Am. Soc. Nephrol.* **33**, 883–887 (2022).

291. *primer_schemes/nCoV-2019 at Master · Artic-Network/artic-ncov2019*. (Github).

292. Kondili, L. A. *et al.* Impact of the COVID-19 pandemic on hepatitis B and C elimination: An EASL survey. *JHEP Rep.* **4**, 100531 (2022).

293. Peto, T. & UK COVID-19 Lateral Flow Oversight Team. COVID-19: Rapid antigen detection for SARS-CoV-2 by lateral flow assay: A national systematic evaluation of sensitivity and specificity for mass-testing. *EClinicalMedicine* **36**, 100924 (2021).

294. Lumley, S. F. *et al.* The duration, dynamics and determinants of SARS-CoV-2 antibody responses in individual healthcare workers. *Clin. Infect. Dis.* (2021) doi:10.1093/cid/ciab004.

295. Lumley, S. F. *et al.* Antibody Status and Incidence of SARS-CoV-2 Infection in Health

Care Workers. *N. Engl. J. Med.* (2020) doi:10.1056/NEJMoa2034545.

296. Eyre, D. W. *et al.* Quantitative SARS-CoV-2 anti-spike responses to Pfizer-BioNTech and Oxford-AstraZeneca vaccines by previous infection status. *Clin. Microbiol. Infect.* **27**, 1516.e7–1516.e14 (2021).

297. Lumley, S. F. *et al.* An observational cohort study on the incidence of severe acute respiratory syndrome Coronavirus 2 (SARS-CoV-2) infection and B.1.1.7 variant infection in healthcare workers by antibody and vaccination status. *Clin. Infect. Dis.* **74**, 1208–1219 (2022).

298. McNaughton, A. L. *et al.* Fatal COVID-19 outcomes are associated with an antibody response targeting epitopes shared with endemic coronaviruses. *JCI Insight* **7**, (2022).

299. Page, M. J. *et al.* The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ* **372**, n71 (2021).

300. Bakhshizadeh, F. *et al.* Efficacy of tenofovir disoproxil fumarate therapy in nucleoside-analogue naive Iranian patients treated for chronic hepatitis B. *Hepat. Mon.* **15**, e25749 (2015).

301. Berg, T. *et al.* Long-term efficacy and safety of emtricitabine plus tenofovir DF vs. tenofovir DF monotherapy in adefovir-experienced chronic hepatitis B patients. *J. Hepatol.* **60**, 715–722 (2014).

302. Boyd, A. *et al.* Persistent viremia in human immunodeficiency virus/hepatitis B coinfected patients undergoing long-term tenofovir: virological and clinical implications. *Hepatology* **60**, 497–507 (2014).

303. Boyd, A. *et al.* Low risk of lamivudine-resistant HBV and hepatic flares in treated HIV-HBV-coinfected patients from Côte d'Ivoire. *Antivir. Ther.* **20**, 643–654 (2015).

304. Buti, M. *et al.* Seven-year efficacy and safety of treatment with tenofovir disoproxil fumarate for chronic hepatitis B virus infection. *Dig. Dis. Sci.* **60**, 1457–1464 (2015).

305. Cathcart, A. L. *et al.* No Resistance to Tenofovir Alafenamide Detected through 96

Weeks of Treatment in Patients with Chronic Hepatitis B Infection. *Antimicrob. Agents Chemother.* **62**, (2018).

306. Chang, T.-T. *et al.* A dose-ranging study of the efficacy and tolerability of entecavir in Lamivudine-refractory chronic hepatitis B patients. *Gastroenterology* **129**, 1198–1209 (2005).

307. Chang, T.-T. *et al.* Results of up to 2 years of entecavir vs lamivudine therapy in nucleoside-naïve HBeAg-positive patients with chronic hepatitis B. *J. Viral Hepat.* **16**, 784–789 (2009).

308. Chen, E. Q. *et al.* A comparison of treatment with adefovir and entecavir for chronic hepatitis B in China: The 2-year results of a prospective study: Adefovir versus Entercavir for Chronic Hepatitis B. *Hepat. Mon.* **11**, 27–31 (2011).

309. Cho, J.-Y. *et al.* Long-term real-world entecavir therapy in treatment-naïve hepatitis B patients: base-line hepatitis B virus DNA and hepatitis B surface antigen levels predict virologic response. *Korean J. Intern. Med.* **32**, 636–646 (2017).

310. Corsa, A. C. *et al.* No resistance to tenofovir disoproxil fumarate through 96 weeks of treatment in patients with lamivudine-resistant chronic hepatitis B. *Clin. Gastroenterol. Hepatol.* **12**, 2106–12.e1 (2014).

311. Deng, X.-L., Li, Q.-L. & Guo, J.-J. Dynamics of lamivudine-resistant hepatitis B virus strains in patients with entecavir rescue therapy. *Virus Genes* **47**, 1–9 (2013).

312. Fung, S. *et al.* Randomized comparison of tenofovir disoproxil fumarate vs emtricitabine and tenofovir disoproxil fumarate in patients with lamivudine-resistant chronic hepatitis B. *Gastroenterology* **146**, 980–988 (2014).

313. Gish, R. G. *et al.* Entecavir therapy for up to 96 weeks in patients with HBeAg-positive chronic hepatitis B. *Gastroenterology* **133**, 1437–1444 (2007).

314. Gwak, G.-Y. *et al.* A comparison of clevudine and entecavir for treatment-naïve patients with chronic hepatitis B: results after 2 years of treatment. *Hepatol. Int.* **7**, 106–110

(2013).

315. Ha, N. B. *et al.* Medication nonadherence with long-term management of patients with hepatitis B e antigen-negative chronic hepatitis B. *Dig. Dis. Sci.* **56**, 2423–2431 (2011).

316. Heathcote, E. J. *et al.* Three-year efficacy and safety of tenofovir disoproxil fumarate treatment for chronic hepatitis B. *Gastroenterology* **140**, 132–143 (2011).

317. Kamezaki, H. *et al.* Emergence of entecavir-resistant mutations in nucleos(t)ide-naive Japanese patients infected with hepatitis B virus: virological breakthrough is also dependent on adherence to medication. *Scand. J. Gastroenterol.* **46**, 1111–1117 (2011).

318. Kamezaki, H. *et al.* Adherence to medication is a more important contributor to viral breakthrough in chronic hepatitis B patients treated with entecavir than in those with Lamivudine. *Int. J. Med. Sci.* **10**, 567–574 (2013).

319. Karino, Y. *et al.* Efficacy and resistance of entecavir following 3 years of treatment of Japanese patients with lamivudine-refractory chronic hepatitis B. *Hepatol. Int.* **4**, 414–422 (2010).

320. Kim, H. J. *et al.* Comparison between clevudine and entecavir treatment for antiviral-naïve patients with chronic hepatitis B. *Liver Int.* **30**, 834–840 (2010).

321. Kim, H. J. *et al.* Rescue therapy for lamivudine-resistant chronic hepatitis B: comparison between entecavir 1.0 mg monotherapy, adefovir monotherapy and adefovir add-on lamivudine combination therapy. *J. Gastroenterol. Hepatol.* **25**, 1374–1380 (2010).

322. Liang, X. *et al.* Long-term efficacy and safety of tenofovir disoproxil fumarate in Chinese patients with chronic hepatitis B: 5-year results. *Hepatol. Int.* **13**, 260–269 (2019).

323. Lim, Y.-S. *et al.* Tenofovir monotherapy versus tenofovir and entecavir combination therapy in patients with entecavir-resistant chronic hepatitis B with multiple drug failure: results of a randomised trial. *Gut* **65**, 852–860 (2016).

324. Lim, Y.-S. *et al.* Tenofovir monotherapy versus tenofovir and entecavir combination therapy in adefovir-resistant chronic hepatitis B patients with multiple drug failure: results

of a randomised trial. *Gut* **65**, 1042–1051 (2016).

325. Liu, Y. *et al.* No detectable resistance to tenofovir disoproxil fumarate in HBeAg+ and HBeAg- patients with chronic hepatitis B after 8 years of treatment. *J. Viral Hepat.* **24**, 68–74 (2017).

326. Marcellin, P. *et al.* Tenofovir disoproxil fumarate versus adefovir dipivoxil for chronic hepatitis B. *N. Engl. J. Med.* **359**, 2442–2455 (2008).

327. Pan, C. Q. *et al.* Efficacy and safety of tenofovir disoproxil fumarate in Asian-Americans with chronic hepatitis B in community settings. *PLoS One* **9**, e89789 (2014).

328. Sherman, M. *et al.* Entecavir for treatment of lamivudine-refractory, HBeAg-positive chronic hepatitis B. *Gastroenterology* **130**, 2039–2049 (2006).

329. Shin, S. R. *et al.* A comparison of 48-week treatment efficacy between clevudine and entecavir in treatment-naïve patients with chronic hepatitis B. *Hepatol. Int.* **5**, 664–670 (2011).

330. Singla, B. *et al.* Response to potent anti-HBV agents in chronic hepatitis B and combined effect of HBV reverse transcriptase mutations. *Gene* **567**, 22–30 (2015).

331. Snow-Lampart, A. *et al.* No resistance to tenofovir disoproxil fumarate detected after up to 144 weeks of therapy in patients monoinfected with chronic hepatitis B virus. *Hepatology* **53**, 763–773 (2011).

332. Stephan, C. *et al.* Impact of tenofovir-containing antiretroviral therapy on chronic hepatitis B in a cohort co-infected with human immunodeficiency virus. *J. Antimicrob. Chemother.* **56**, 1087–1093 (2005).

333. Suzuki, F. *et al.* Efficacy and safety of entecavir in lamivudine-refractory patients with chronic hepatitis B: randomized controlled trial in Japanese patients. *J. Gastroenterol. Hepatol.* **23**, 1320–1326 (2008).

334. Suzuki, F. *et al.* Long-term outcome of entecavir treatment of nucleos(t)ide analogue-naïve chronic hepatitis B patients in Japan. *J. Gastroenterol.* **54**, 182–193

(2019).

335. Tenney, D. J. *et al.* Two-year assessment of entecavir resistance in Lamivudine-refractory hepatitis B virus patients reveals different clinical outcomes depending on the resistance substitutions present. *Antimicrob. Agents Chemother.* **51**, 902–911 (2007).

336. Tsai, M.-C. *et al.* A comparison of telbivudine and entecavir for chronic hepatitis B in real-world clinical practice. *J. Antimicrob. Chemother.* **67**, 696–699 (2012).

337. Yang, D.-H. *et al.* Tenofovir disoproxil fumarate is superior to lamivudine plus adefovir in lamivudine-resistant chronic hepatitis B patients. *World J. Gastroenterol.* **21**, 2746–2753 (2015).

338. Yokosuka, O. *et al.* Long-term use of entecavir in nucleoside-naïve Japanese patients with chronic hepatitis B infection. *J. Hepatol.* **52**, 791–799 (2010).

339. Yuen, M.-F. *et al.* Two-year treatment outcome of chronic hepatitis B infection treated with besifovir vs. entecavir: results from a multicentre study. *J. Hepatol.* **62**, 526–532 (2015).

340. Zhou, J. *et al.* Efficacy and Safety of Tenofovir Disoproxil Treatment for Chronic Hepatitis B Patients with Genotypic Resistance to Other Nucleoside Analogues: A Prospective Study. *Chin. Med. J.* **130**, 914–919 (2017).