Title Page (including author details, affiliations and word count)

Title: Nanopore sequencing from extraction-free direct PCR of dried serum spots for portable hepatitis B virus drug-resistance typing

Running title: Sequencing of hepatitis B virus directly from dried serum spot

Stuart Astbury[a,b,c], Marcia Maria Da Costa Nunes Soares[d], Emmanuel Peprah[e], Barnabas King[b,e], Ana Carolina Gomes Jardim[f], Jacqueline Farinha Shimizu[f], Paywast Jalal[g], Chiman H Saeed[h], Furat T Sabeer[i], William L Irving[b,c,e], Alexander W Tarr[b,c,e*], C Patrick McClure[b,e]

[a] Nottingham Digestive Diseases Centre, School of Medicine, University of Nottingham, UK

[b] NIHR Nottingham Biomedical Research Centre, Nottingham University Hospitals NHS Trust and the University of Nottingham, UK

[c] MRC/EPSRC Nottingham Molecular Pathology Node, University of Nottingham, UK

[d] Instituto Adolfo Lutz, Brazilian Ministry of Health, São Paulo, Brazil

[e] School of Life Sciences, University of Nottingham, UK

[f] Institute of Biomedical Sciences, Federal University of Uberlândia, Uberlândia, Brazil.

[g] Biology Department, Faculty of Science, University of Sulaimani, Sulaymaniyah, Iraq

[h] Medical Research Center, Hawler Medical University, Erbil, Iraq

[i] Central Public Health Laboratory, Erbil, Iraq

*Corresponding author: alex.tarr@nottingham.ac.uk

Abstract word count: 250

Main text word count: 2482

Keywords Dried blood spot; nanopore sequencing; portable diagnostic test; hepatitis B virus; resistance-associated substitutions; genotyping

*Abstract

Abstract

Background: Effective drug regimens for the treatment of hepatitis B virus (HBV) infections are essential to achieve the World Health Organisation commitment to eliminate viral hepatitis by 2030. Lamivudine (3TC) is widely used in countries with high levels of chronic HBV, however resistance has been shown to occur in up to 50% of individuals receiving continuous monotherapy for 4 years. Telbivudine (LdT) is now more commonly used in place of lamivudine but is ineffective against 3TC-resistant HBV. Genotyping and identification of resistance-associated substitutions (RAS) is not practical in many locations.

Objectives: A novel assay was designed to enable HBV genotyping and characterisation of resistance mutations directly from serum samples stored on filter paper, using Sanger and MinION sequencing.

Study design: The assay was applied to a cohort of 30 samples stored on filter paper for several years with HBV viral loads ranging from $8.2x10^8$ to 635 IU/mL. A set of 6 high-titre samples were used in a proof-of-principle study using the MinION sequencer.

Results: The assay allowed determination of HBV genotype and elucidation of RAS down to 600 IU/mL using a 550bp amplicon. Sequencing of a 1.2kb amplicon using a MinION sequencer gave results consistent with Sanger and allowed the identification of minor populations of variants.

Conclusions: We present two methods for reliable HBV sequencing and RAS identification using methods suitable for resource-limited environments. This is the first demonstration of extraction-free DNA sequencing direct from DSS using MinION and these workflows are adaptable to the investigation of other DNA viruses.

***Highlights (for review)**

- Genotyping and screening for resistance associated substitutions in hepatitis B virus genomes is possible using an extraction-free method with dried serum spots (DSS) and direct PCR enzymes.
- Longer fragments can be recovered from high-titre samples and are suitable for deep sequencing using the Oxford Nanopore MinION.
- This method allows effective HBV diagnostic testing to be carried out in resource-limited settings.

Background and objectives

Hepatitis B virus (HBV) currently infects an estimated 257 million people worldwide and there is an urgent need for screening and surveillance tools to assess HBV in low and middle-income countries [1]. In some African regions it is estimated that 6% of the population are infected and only one in every ten children is vaccinated [2, 3]. HBV has a complex and error-prone replication cycle and there are many well-characterised mutations across the HBV genome, conferring resistance to therapy, an increase in replication efficiency (polymerase), immune- and diagnostic test-escape (S/pre-core), or increased pathogenicity (reviewed in [4]).

Treatment of chronic HBV typically utilises long term monotherapy with a polymerase inhibitor. Resistance mutations to the polymerase inhibitor lamivudine (3TC) are well documented, occurring in 51 % of patients receiving monotherapy for 4 years [5]. Telbivudine (LdT), which is becoming increasingly more prescribed, is ineffective against 3TC-resistant HBV. Although lamivudine and telbivudine are non-preferred treatment options, as defined by the American Society for the Study of Liver Diseases (AASLD) and the European Association for the Study of the Liver (EASL) [6, 7], they remain widely used in resource-limited countries with high levels of chronic HBV. Furthermore, while tenofovir (TDF) is the preferred option, 3TC and LdT have been suggested as cost-effective treatments during pregnancy to limit vertical transmission [8, 9].

An added complication in HBV therapy is the use of 3TC for the treatment of HIV. 3TC currently forms part of large -scale anti-HIV programs in parts of Africa, however HBV prevalence and the impact of HIV treatment on HBV resistance associated substitutions (RAS) is not effectively monitored [10, 11].

A primary biomarker informing clinical management of chronic HBV is the serum level of hepatitis B surface antigen (HBsAg). Several mutations are known to lead to low or false-negative results in diagnostic quantification assays (reviewed in [12]), but these mutations are not routinely investigated. Greater understanding of diagnostic escape mutations will inform improved patient management.

RAS-typing and genotyping can be achieved by several methods. Sanger sequencing is the gold standard in clinical applications and represents the most accessible and affordable choice globally, with some countries able to access overnight sequencing from room temperature-shipped PCR products. Additionally, the introduction of third-generation sequencing platforms such as the Oxford Nanopore Technologies (ONT) MinION enables sequencing with no theoretical upper limit on read length, enabling sequencing of entire genes or viral genomes in a single read. The MinION sequencer is extremely portable, being powered through a laptop, and has been used effectively during the Ebola outbreak in West Africa [13] and in tracking the spread of Zika virus in Brazil [14]. Field application of the MinION platform has been enhanced with recent advances including improved R9.4 flow cells with increased accuracy and software such as Nanopolish [15], which works with signal-level data from the sequencer allowing generation of more accurate consensus sequences. More recently, a methodology for the MinION platform has been developed and applied to HBV sequencing and haplotyping which utilises the circular nature of the genome to generate concatenated single genome replicates [16]. While this work marked a significant technical innovation, it has limited application to diverse clinical samples due to the high viral load required for the sample (>10^8 IU/ml), as therapy is recommended when serum HBV DNA is as low as >2,000 IU/ml (and other criteria are met) [6, 7].

Here we combined both Sanger and MinION sequencing with dried serum spot (DSS) sampling and extraction-free direct PCR to develop proof-of-principle workflows for generating clinically relevant HBV sequence data in regions without access to conventional sample storage or a cold chain. A range of primers were developed and assessed targeting the overlapping ORFs containing S and the reverse transcriptase (RT) domain of the polymerase gene, which facilitated genotyping and RAS-typing. As the HBV genome is a DNA molecule it allows a clinical sample to be added directly to a PCR mix with no prior extraction. While this work was conducted with dried serum spots, it would be compatible with dried blood spot samples and a range of pathogens with DNA genomes.

Study design

Samples: All HBV DNA-positive samples were surplus material obtained for routine diagnostics. A Brazilian cohort obtained from eight centres in São Paulo State were sampled between July 2016 and April 2017. Virus titre was determined by RealTime HBV Amplification kit (Abbott). Iraqi samples were collected and processed at the Erbil Central Laboratory in 2017 and viral load was determined by Artus HBV PCR kit (Qiagen). Sera were stored at -20°C before preparation of DSS cards. All samples were obtained for routine diagnostic investigation of HBV, and collection of surplus material from this consented activity was locally approved by both institutes (Adolfo Lutz Institute, Brazil; University of Sulaimani, Iraq) for the extended diagnostic development presented in this study.

Sample preparation: 25-30 µL serum was spotted onto a Whatman® Protein Saver™ 903 Card (GE Healthcare), saturating the 12 mm diameter area demarcated on the DSS cards. Cards were air-dried at room temperature for ~2 h and stored at 4°C with Silica gel sachets. Samples were stored under these conditions for up to 2 years before the work presented

here was carried out. Reactions were prepared using a hole punch with a 3 mm diameter, each filter paper punch representing ~1.5 µL of serum. To prevent cross-contamination a punch of clean filter paper was performed between each sample. A clean filter paper (punched directly after the final sample) was included in each PCR run as a negative control. To determine the level of carryover using this protocol, reactions were set up using the most sensitive assay (F1/R1 primers, table 1), and 9 high titre samples (ranging from 10^9 to 10^6 IU/mL) were intermingled with samples with low or undetectable viral loads, with no amplification observed in samples with low or negligible HBV titre .

PCR: Primers are shown in table 1. Phusion Blood Direct (Thermo Fisher) was used in 25 µL reactions (12.5 µL 2x Phusion Blood Direct mastermix, 0.5 µM of each primer). 3 mm punches taken from DSS cards were then added directly to each tube. PCRs were performed as follows: 98 °C for 5 minutes, 55 X [98 °C for 1 second, 50 °C for 5 seconds and 72 °C for 20 seconds/kb], final extension 72 °C for 1 minute.

Sanger sequencing: PCR products were diluted 1:10 with nuclease-free water for Sanger sequencing (Source Bioscience, Nottingham, UK) using primers F1+R1 or F2+R2 for short and long amplicons, respectively. Reads were assembled into contiguous sequences and aligned using MUSCLE in MEGA X [17, 18].

MinION sequencing library preparation: MinION-barcoded samples were prepared using a two stage PCR and SQK-LWB001 library preparation kit (ONT). First round PCR was carried out as above, using 25 µL Hemo KlenTaq reactions with 3 mm punches from DSS and primers F3 and R3 (table 1). 1 µL of reaction product was used as template for a 50 µl PCR: 25 µL 2x LongAmp HotStart Taq master mix, 1.5 µL of barcoded primer mix (ONT) and 22.5 µL nuclease free water. Purified PCR products were quantified by Qubit using the dsDNA

high sensitivity kit (Thermo Fisher). Based on the method developed by Quick *et al*. [14], amplicons were pooled to achieve a total of 0.3 pM input DNA per MinION flow cell (~260 ng total DNA for a 1.3 kb amplicon). Following adapter ligation, the library was run for 48 hrs on a R9.4.1 flowcell on a MinION Mk II controlled through MinKNOW 1.10.16 software followed by base-calling using Albacore 2.2.2. Adapter-trimmed sequences were uploaded to the NCBI Sequence Read Archive under project ID PRJNA521740.

MinION sequencing analysis: Basecalled reads were trimmed using Porechop 0.2.3 using high stringency settings (--discard_middle and --require_two_barcodes) and retained when Porechop and Albacore barcode aligners agreed. NanoPlot was used to inspect read quality and length, and reads were filtered based on length (min.: 1200; max.: 1300) using NanoFilt [19]. Consensus sequences were assembled *de novo* from processed reads using Canu (v1.7.1) [20]. The same reads were subsequently aligned using Minimap2 [21] to their respective Canu consensus sequences. This alignment was used as the input for further processing in Nanopolish v0.10.2 [15], using the Variants module and the --fix-homopolymers function to generate a corrected consensus sequence. Corrected nanopore consensus sequences were aligned using MUSCLE within MEGA X to their respective Sanger contigs to assess sequence similarity. A full description of the bioinformatics workflow used is included as supplementary data.

Genotyping and variant calling: Genotypes were determined using the web-based tools HBV geno2pheno [22] and HBVseq [23] following removal of primer sequences. For samples sequenced by MinION potential intra-host variants were screened for by aligning processed, filtered reads to the corresponding Sanger sequence. This alignment was then used as the input for LoFreq [24]. Considering the relatively high error rate of individual nanopore reads,

variants below a 10% threshold were filtered out, and variants with significant strand bias were filtered using the default LoFreq settings. Primer sites were masked from the alignments used for variant calling.

Results

Sanger sequencing direct from DSS and genotyping

Initially, primer pair F1/R1 was used with Phusion Blood Direct polymerase for the analysis of 30 HBV-positive serum samples of defined viral load isolated in Brazil. Amplicons were achieved for all but 3 samples using these primers, facilitating genotyping (table 2). Of the 27 amplified samples, genotype was distributed as follows: A1 n=7, A2 n=3, B1 n=1, D1 n=1, D2 n=2, D3 n=8, F1 n=1, F2 n=3. A single sample, Br28, was classed as genotype A but did not match any subtype sequences in geno2pheno.

RAS characterisation

The amplicon produced from the F2/R2 primers permitted Sanger sequencing across aa 169 to aa 250 of the reverse transcriptase (RT) domain of the Pol gene (figure 1, highlighted in red), which is the region critical to identification of RAS. Of those samples for which an amplicon was obtained, 5/25 contained known polymerase RAS (table 2). A further 4 samples contained minor peaks on Sanger traces at known RAS sites (table 2).

MinION sequencing direct from DSS

Having obtained clinically relevant sequence data by Sanger sequencing from DSS, we investigated whether comparable data could be obtained using MinION sequencing in a small proof-of-principle study. Three high titre samples were selected from the Brazilian cohort (Br1, Br2, Br3), along with three samples from a separate cohort sourced from the Kurdistan region of Iraq. The samples sourced from Iraq all had a viral load $>10^8$ IU/mL. All samples were successfully amplified and Sanger sequenced using F1/R1 and F2/R2 primer pairs initially. An attempt was made to amplify whole HBV genomes from DSS using our

direct PCR protocol and a previously published primer set [25], however this was unsuccessful regardless of viral load.

Amplicons of 1,274 bp were successfully generated using primers F3 and R3. A total of 185,349 raw reads were obtained for the 6 samples, ranging from 23869 to 41392 per barcode. Following adapter trimming and further filtering of erroneously long and short reads these counts ranged from 7676 to 11832 per barcode (table 3). A summary of raw sequencing reads acquired over time, and average quality score per read over time is included in figure 2.

De novo consensus sequence building

Consensus sequences were assembled using Canu. Following initial assembly, the only errors observed were single base deletions within homopolymers when compared to Sanger sequences, ranging from 5 to 12 deletions (data not shown). Following a single round of Nanopolish processing all consensus sequences were identical to their Sanger counterparts (figure 3).

Detection of minor variants

Several putative minor variants were detected following alignment of MinION reads with the *de novo* consensus sequence. The majority of these were filtered out by LoFreq. A single variant was detected in sample Iq3, an A > G switch in 19% of reads coding for N155D within the spacer region of polymerase. Inspection of the Sanger trace confirmed the presence of this minor variant.

Discussion

Methods to both genotype and characterise therapeutic and diagnostic mutations can be technically, logistically or financially challenging in regions of the world with a high prevalence of chronic HBV infection. We developed two workflows for generating sequencing data from DSS using Sanger and MinION platforms, obtaining fully analysed sequences in less than 48 hours for both methods from receipt of samples. Deep sequencing data were successfully generated for all six of the samples tested and, following *in silico* processing, produced consensus sequences that were identical to the sequences generated by Sanger sequencing.

While both sequencing methods involve specific expenses, once established they are cost-effective, require no cold chain, and the PCR achieved a lower limit of detection of approximately 600 IU/mL for small amplicons. This level of sensitivity makes both workflows viable options for informing clinical decisions as the recommended limit set for initiating therapy is 2,000 IU/mL [6, 7]. However, while the direct PCR method used here worked well for partial gene amplicons at low viral loads, it was not successful when applied to previously published primer sets for amplifying the entire HBV genome [25] (data not shown), suggesting traditional extraction methods may be preferable for larger (>1.5kb) amplicons.

Several RAS detected in this cohort are clinically significant. The Y100C mutant in the S region of HBsAg has been linked to false negative/low HBsAg tests interpreted as occult infection [26]. This variant was unusually common in our Brazilian cohort, being present in 5/10 genotype A samples, compared to only 2.8% of genotype A samples analysed in a separate study, from South Africa [27].

Our PCR approach specifically targeted the RT domain of Pol to characterise drug resistance mutants. This is particularly important for application in countries where drugs with a low barrier to resistance are commonly prescribed. The V173L, L180M and M204V RAS detected in the polymerase gene indicated resistance to the nucleoside analogues lamivudine and entecavir [28]. The N236S mutation has not been previously described, but other mutations at this site have been shown to lead to breakthrough on adefovir [29] and decrease the efficacy of tenofovir *in vitro* [30]. Further research is required to determine if N236S produces a similar phenotype. Finally M250I, present in one sample as a heterozygous peak in Sanger sequencing, has been shown to lead to lamivudine and telbivudine resistance [28].

Complex selection pressures on HBV lead to evolution of intra-host subpopulations [31]. Accurately determining minor variants and haplotypes is potentially of great clinical value, particularly for a virus requiring long-term therapy allowing the expansion of multi-drug resistant haplotypes. By applying a conservative variant calling approach we identified one minor variant sequence using the MinION workflow (N155D in sample Iq3). This variant was subsequently confirmed by visual analysis of the Sanger read. As residue 155 is located in the spacer region of polymerase it is unlikely to be clinically significant but highlights the potential for detecting low-level variants. Multi-drug resistant populations can arise through sequential RAS introduction over time [32]. Long read high-throughput sequencing technologies such as nanopore allow characterisation of the changing population of HBV variants in high titre samples, which has previously required clonal analysis combined with Sanger sequencing.

We acknowledge a number of limitations of this study. Large amplicons were chosen to utilise nanopore sequencing directly from DSS in this proof-of-concept study, but greater sensitivity could be achieved using a PCR-tiling approach of two or more amplicons that are pooled before sequencing (as demonstrated with the much larger Zika virus genome [14]), or by scaling up reaction mixtures and using more DSS per reaction. Identical sequences were also observed in two pairs of high viral load samples chosen for nanopore sequencing. To confirm that this was not an artefact introduced by cross-contamination during processing, additional, independent DSS punches were taken and sequenced following direct PCR amplification from these samples to verify the sequences. Given the low substitution rate of HBV and the fact that these pairs of patients were sampled from the same geographic region it is likely that these are genuinely identical sequences. However, errors in preparing DSS cards cannot be discounted. Although we observed no carryover between high titre samples and those with an undetectable viral load, more stringent decontamination between punches could be considered if this workflow was applied in a clinical setting. We also highlight issues with obtaining amplicons >1.5kb using the direct PCR enzyme, which limits the ability of this direct method to detect linked mutations in long genomes.

In summary, we describe two approaches for rapid genotyping and RAS detection in HBV, using a novel analyte, dried serum spots, which are applicable in resource-limited settings and require little existing infrastructure. The results presented here demonstrate the utility of direct PCR enzymes and DSS together in a clinical context. We have also demonstrated, for the first time, that nanopore sequencing can be applied directly to samples amplified

from DSS, with no requirement for extraction. Reliable sequence data was generated using the MinION sequencer, significantly reducing the requirements for laboratory infrastructure.

Funding

Funding was provided by the Nottingham Molecular Pathology Node (MRC/EPSRC grant MR/N005953/1) and the National Institute for Health Research (NIHR) Nottingham Biomedical Research Centre.

Competing interests

The authors have no competing interests to declare.

References

[1] Lemoine M, Eholie S, Lacombe K. Reducing the neglected burden of viral hepatitis in Africa: strategies for a global approach. J Hepatol. 2015;62:469-76.

[2] Graber-Stiehl I. The silent epidemic killing more people than HIV, malaria or TB. Nature. 2018;564:24-6.

[3] O'Hara GA, McNaughton AL, Maponga T, Jooste P, Ocama P, Chilengi R, et al. Hepatitis B virus infection as a neglected tropical disease. PLoS Negl Trop Dis. 2017;11:e0005842.

[4] Lapinski TW, Pogorzelska J, Flisiak R. HBV mutations and their clinical significance. Adv Med Sci. 2012;57:18-22.

[5] Thompson AJ, Ayres A, Yuen L, Bartholomeusz A, Bowden DS, Iser DM, et al. Lamivudine resistance in patients with chronic hepatitis B: role of clinical and virological factors. J Gastroenterol Hepatol. 2007;22:1078-85.

[6] Liver EAftSoT. EASL 2017 Clinical Practice Guidelines on the management of hepatitis B virus infection. J Hepatol. 2017;67:370-98.

[7] Terrault NA, Lok ASF, McMahon BJ, Chang KM, Hwang JP, Jonas MM, et al. Update on prevention, diagnosis, and treatment of chronic hepatitis B: AASLD 2018 hepatitis B guidance. Hepatology. 2018;67:1560-99.

[8] Han GR, Cao MK, Zhao W, Jiang HX, Wang CM, Bai SF, et al. A prospective and open-label study for the efficacy and safety of telbivudine in pregnancy for the prevention of perinatal transmission of hepatitis B virus infection. J Hepatol. 2011;55:1215-21.

[9] Xu WM, Cui YT, Wang L, Yang H, Liang ZQ, Li XM, et al. Lamivudine in late pregnancy to prevent perinatal transmission of hepatitis B virus infection: a multicentre, randomized, double-blind, placebo-controlled study. J Viral Hepat. 2009;16:94-103.

[10] Barth RE, Huijgen Q, Taljaard J, Hoepelman AI. Hepatitis B/C and HIV in sub-Saharan Africa: an association between highly prevalent infectious diseases. A systematic review and meta-analysis. Int J Infect Dis. 2010;14:e1024-31.

[11] Matthews PC, Geretti AM, Goulder PJ, Klenerman P. Epidemiology and impact of HIV coinfection with hepatitis B and hepatitis C viruses in Sub-Saharan Africa. J Clin Virol. 2014;61:20-33.

[12] Weber B. Genetic variability of the S gene of hepatitis B virus: clinical and diagnostic impact. J Clin Virol. 2005;32:102-12.

[13] Quick J, Loman NJ, Duraffour S, Simpson JT, Severi E, Cowley L, et al. Real-time, portable genome sequencing for Ebola surveillance. Nature. 2016;530:228-32.

[14] Quick J, Grubaugh ND, Pullan ST, Claro IM, Smith AD, Gangavarapu K, et al. Multiplex PCR method for MinION and Illumina sequencing of Zika and other virus genomes directly from clinical samples. Nat Protoc. 2017;12:1261-76.

[15] Loman NJ, Quick J, Simpson JT. A complete bacterial genome assembled de novo using only nanopore sequencing data. Nat Methods. 2015;12:733-5.

[16] McNaughton AL, Roberts HE, Bonsall D, de Cesare M, Mokaya J, Lumley SF, et al. Illumina and Nanopore methods for whole genome sequencing of hepatitis B virus (HBV). Scientific Reports. 2019;9:7081.

[17] Kumar S, Stecher G, Li M, Knyaz C, Tamura K. MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. Mol Biol Evol. 2018;35:1547-9.

[18] Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 2004;32:1792-7.

[19] De Coster W, D'Hert S, Schultz DT, Cruts M, Van Broeckhoven C. NanoPack: visualizing and processing long-read sequencing data. Bioinformatics. 2018;34:2666-9.
[20] Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. Genome Res. 2017;27:722-36.
[21] Li H. Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics. 2018;34:3094-100.
[22] Beggel B, Neumann-Fraune M, Doring M, Lawyer G, Kaiser R, Verheyen J, et al. Genotyping hepatitis B virus dual infections using population-based sequence data. J Gen Virol. 2012;93:1899-907.
[23] Rhee SY, Margeridon-Thermet S, Nguyen MH, Liu TF, Kagan RM, Beggel B, et al. Hepatitis B virus reverse transcriptase sequence variant database for sequence analysis and mutation discovery. Antiviral Res. 2010;88:269-75.
[24] Wilm A, Aw PP, Bertrand D, Yeo GH, Ong SH, Wong CH, et al. LoFreq: a sequence-quality aware, ultra-sensitive variant caller for uncovering cell-population heterogeneity from high-throughput sequencing datasets. Nucleic Acids Res. 2012;40:11189-201.
[25] Gunther S, Li BC, Miska S, Kruger DH, Meisel H, Will H. 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. 1995;69:5437-44.
[26] Martin CM, Welge JA, Rouster SD, Shata MT, Sherman KE, Blackard JT. Mutations associated with occult hepatitis B virus infection result in decreased surface antigen expression in vitro. J Viral Hepat. 2012;19:716-23.
[27] Gencay M, Vermeulen M, Neofytos D, Westergaard G, Pabinger S, Kriegner A, et al. Substantial variation in the hepatitis B surface antigen (HBsAg) in hepatitis B virus (HBV)-positive patients from South Africa: Reliable detection of HBV by the Elecsys HBsAg II assay. J Clin Virol. 2018;101:38-43.
[28] Zoulim F, Locarnini S. Hepatitis B virus resistance to nucleos(t)ide analogues. Gastroenterology. 2009;137:1593-608 e1-2.
[29] Sheldon J, Camino N, Rodes B, Bartholomeusz A, Kuiper M, Tacke F, et al. Selection of hepatitis B virus polymerase mutations in HIV-coinfected patients treated with tenofovir. Antivir Ther. 2005;10:727-34.
[30] Brunelle MN, Jacquard AC, Pichoud C, Durantel D, Carrouee-Durantel S, Villeneuve JP, et al. Susceptibility to antivirals of a human HBV strain with mutations conferring resistance to both lamivudine and adefovir. Hepatology. 2005;41:1391-8.
[31] McNaughton AL, D'Arienzo V, Ansari MA, Lumley SF, Littlejohn M, Revill P, et al. Insights From Deep Sequencing of the HBV Genome-Unique, Tiny, and Misunderstood. Gastroenterology. 2019;156:384-99.
[32] Yim HJ, Hussain M, Liu Y, Wong SN, Fung SK, Lok AS. Evolution of multi-drug resistant hepatitis B virus during sequential therapy. Hepatology. 2006;44:703-12.
[33] Hayer J, Jadeau F, Deléage G, Kay A, Zoulim F, Combet C. HBVdb: a knowledge database for Hepatitis B Virus. Nucleic Acids Research. 2013;41:D566-D70.

Figure captions

Figure 1. Cartoon highlighting the locations of PCR amplicons. Overlapping ORFs in the HBV genome (genotype A reference isolate X02763) are shown in blue. PCR amplicons generated in this study are shown in orange, and the crucial RT region associated with treatment resistance mutations (aa169 – 250) is shown in red. The F1+R1 amplicon is sufficient for genotyping and limited detection of sAg diagnostic escape mutants. The region of the reverse transcriptase (RT) domain in which resistance associated substitutions (RAS) arise (aa169 – 250), shown in red, is encompassed by the F2+R2 amplicon.

Figure 2. Metrics from MinION sequencing run. Both plots were generated from raw reads assigned barcodes by Albacore without further filtering. Quality scores are standard Phred scores produced by Albacore during basecalling, data is presented as mean Phred score per read with min and max.

Figure 3: Maximum likelihood tree comparing sequences sourced from direct PCR using both Sanger and nanopore sequencing methods. Nanopore data was first analysed using Nanopolish, before comparing to Sanger sequences. Sequences were generated using the amplicons generated from the F3/R3 primers. The ML tree was inferred using a general time reversible model within MEGA X [17]. Statistical robustness was assessed using bootstrap resampling of 1,000 pseudoreplicates. The tree with the highest log-likelihood is shown.

359

Background and objectives

Hepatitis B virus (HBV) currently infects an estimated 257 million people worldwide and there is an urgent need for screening and surveillance tools to assess HBV in low and middle-income countries [1]. In some African regions it is estimated that 6% of the population are infected and only one in every ten children is vaccinated [2, 3]. HBV has a complex and error-prone replication cycle and there are many well-characterised mutations across the HBV genome, conferring resistance to therapy, an increase in replication efficiency (polymerase), immune- and diagnostic test-escape (S/pre-core), or increased pathogenicity (reviewed in [4]).

Treatment of chronic HBV typically utilises long term monotherapy with a polymerase inhibitor. Resistance mutations to the polymerase inhibitor lamivudine (3TC) are well documented, occurring in 51 % of patients receiving monotherapy for 4 years [5]. Telbivudine (LdT), which is becoming increasingly more prescribed, is ineffective against 3TC-resistant HBV. Although lamivudine and telbivudine are non-preferred treatment options, as defined by the American Society for the Study of Liver Diseases (AASLD) and the European Association for the Study of the Liver (EASL) [6, 7], they remain widely used in resource-limited countries with high levels of chronic HBV. Furthermore, while tenofovir (TDF) is the preferred option, 3TC and LdT have been suggested as cost-effective treatments during pregnancy to limit vertical transmission [8, 9].

An added complication in HBV therapy is the use of 3TC for the treatment of HIV. 3TC currently forms part of large -scale anti-HIV programs in parts of Africa, however HBV prevalence and the impact of HIV treatment on HBV resistance associated substitutions (RAS) is not effectively monitored [10, 11].

A primary biomarker informing clinical management of chronic HBV is the serum level of hepatitis B surface antigen (HBsAg). Several mutations are known to lead to low or false-negative results in diagnostic quantification assays (reviewed in [12]), but these mutations are not routinely investigated. Greater understanding of diagnostic escape mutations will inform improved patient management.

RAS-typing and genotyping can be achieved by several methods. Sanger sequencing is the gold standard in clinical applications and represents the most accessible and affordable choice globally, with some countries able to access overnight sequencing from room temperature-shipped PCR products. Additionally, the introduction of third-generation sequencing platforms such as the Oxford Nanopore Technologies (ONT) MinION enables sequencing with no theoretical upper limit on read length, enabling sequencing of entire genes or viral genomes in a single read. The MinION sequencer is extremely portable, being powered through a laptop, and has been used effectively during the Ebola outbreak in West Africa [13] and in tracking the spread of Zika virus in Brazil [14]. Field application of the MinION platform has been enhanced with recent advances including improved R9.4 flow cells with increased accuracy and software such as Nanopolish [15], which works with signal-level data from the sequencer allowing generation of more accurate consensus sequences. More recently, a methodology for the MinION platform has been developed and applied to HBV sequencing and haplotyping which utilises the circular nature of the genome to generate concatenated single genome replicates [16]. While this work marked a significant technical innovation, it has limited application to diverse clinical samples due to the high viral load required for the sample (>10^8 IU/ml), as therapy is recommended when serum HBV DNA is as low as >2,000 IU/ml (and other criteria are met) [6, 7].

Here we combined both Sanger and MinION sequencing with dried serum spot (DSS) sampling and extraction-free direct PCR to develop proof-of-principle workflows for generating clinically relevant HBV sequence data in regions without access to conventional sample storage or a cold chain. A range of primers were developed and assessed targeting the overlapping ORFs containing S and the reverse transcriptase (RT) domain of the polymerase gene, which facilitated genotyping and RAS-typing. As the HBV genome is a DNA molecule it allows a clinical sample to be added directly to a PCR mix with no prior extraction. While this work was conducted with dried serum spots, it would be compatible with dried blood spot samples and a range of pathogens with DNA genomes.

Study design

Samples: All HBV DNA-positive samples were surplus material obtained for routine diagnostics. A Brazilian cohort obtained from eight centres in São Paulo State were sampled between July 2016 and April 2017. Virus titre was determined by RealTime HBV Amplification kit (Abbott). Iraqi samples were collected and processed at the Erbil Central Laboratory in 2017 and viral load was determined by Artus HBV PCR kit (Qiagen). Sera were stored at -20°C before preparation of DSS cards. All samples were obtained for routine diagnostic investigation of HBV, and collection of surplus material from this consented activity was locally approved by both institutes (Adolfo Lutz Institute, Brazil; University of Sulaimani, Iraq) for the extended diagnostic development presented in this study.

Sample preparation: 25-30 µL serum was spotted onto a Whatman® Protein Saver™ 903 Card (GE Healthcare), saturating the 12 mm diameter area demarcated on the DSS cards. Cards were air-dried at room temperature for ~2 h and stored at 4°C with Silica gel sachets. Samples were stored under these conditions for up to 2 years before the work presented

here was carried out. Reactions were prepared using a hole punch with a 3 mm diameter, each filter paper punch representing ~1.5 µL of serum. To prevent cross-contamination a punch of clean filter paper was performed between each sample. A clean filter paper (punched directly after the final sample) was included in each PCR run as a negative control. To determine the level of carryover using this protocol, reactions were set up using the most sensitive assay (F1/R1 primers, table 1), and 9 high titre samples (ranging from 10^9 to 10^6 IU/mL) were intermingled with samples with low or undetectable viral loads, with no amplification observed in samples with low or negligible HBV titre .

PCR: Primers are shown in table 1. Phusion Blood Direct (Thermo Fisher) was used in 25 µL reactions (12.5 µL 2x Phusion Blood Direct mastermix, 0.5 µM of each primer). 3 mm punches taken from DSS cards were then added directly to each tube. PCRs were performed as follows: 98 °C for 5 minutes, 55 X [98 °C for 1 second, 50 °C for 5 seconds and 72 °C for 20 seconds/kb], final extension 72 °C for 1 minute.

Sanger sequencing: PCR products were diluted 1:10 with nuclease-free water for Sanger sequencing (Source Bioscience, Nottingham, UK) using primers F1+R1 or F2+R2 for short and long amplicons, respectively. Reads were assembled into contiguous sequences and aligned using MUSCLE in MEGA X [17, 18].

MinION sequencing library preparation: MinION-barcoded samples were prepared using a two stage PCR and SQK-LWB001 library preparation kit (ONT). First round PCR was carried out as above, using 25 µL Hemo KlenTaq reactions with 3 mm punches from DSS and primers F3 and R3 (table 1). 1 µL of reaction product was used as template for a 50 µl PCR: 25 µL 2x LongAmp HotStart Taq master mix, 1.5 µL of barcoded primer mix (ONT) and 22.5 µL nuclease free water. Purified PCR products were quantified by Qubit using the dsDNA

high sensitivity kit (Thermo Fisher). Based on the method developed by Quick *et al*. [14], amplicons were pooled to achieve a total of 0.3 pM input DNA per MinION flow cell (~260 ng total DNA for a 1.3 kb amplicon). Following adapter ligation, the library was run for 48 hrs on a R9.4.1 flowcell on a MinION Mk II controlled through MinKNOW 1.10.16 software followed by base-calling using Albacore 2.2.2. Adapter-trimmed sequences were uploaded to the NCBI Sequence Read Archive under project ID PRJNA521740.

MinION sequencing analysis: Basecalled reads were trimmed using Porechop 0.2.3 using high stringency settings (--discard_middle and --require_two_barcodes) and retained when Porechop and Albacore barcode aligners agreed. NanoPlot was used to inspect read quality and length, and reads were filtered based on length (min.: 1200; max.: 1300) using NanoFilt [19]. Consensus sequences were assembled *de novo* from processed reads using Canu (v1.7.1) [20]. The same reads were subsequently aligned using Minimap2 [21] to their respective Canu consensus sequences. This alignment was used as the input for further processing in Nanopolish v0.10.2 [15], using the Variants module and the --fix-homopolymers function to generate a corrected consensus sequence. Corrected nanopore consensus sequences were aligned using MUSCLE within MEGA X to their respective Sanger contigs to assess sequence similarity. A full description of the bioinformatics workflow used is included as supplementary data.

Genotyping and variant calling: Genotypes were determined using the web-based tools HBV geno2pheno [22] and HBVseq [23] following removal of primer sequences. For samples sequenced by MinION potential intra-host variants were screened for by aligning processed, filtered reads to the corresponding Sanger sequence. This alignment was then used as the input for LoFreq [24]. Considering the relatively high error rate of individual nanopore reads,

variants below a 10% threshold were filtered out, and variants with significant strand bias were filtered using the default LoFreq settings. Primer sites were masked from the alignments used for variant calling.

Results

Sanger sequencing direct from DSS and genotyping

Initially, primer pair F1/R1 was used with Phusion Blood Direct polymerase for the analysis of 30 HBV-positive serum samples of defined viral load isolated in Brazil. Amplicons were achieved for all but 3 samples using these primers, facilitating genotyping (table 2). Of the 27 amplified samples, genotype was distributed as follows: A1 n=7, A2 n=3, B1 n=1, D1 n=1, D2 n=2, D3 n=8, F1 n=1, F2 n=3. A single sample, Br28, was classed as genotype A but did not match any subtype sequences in geno2pheno.

RAS characterisation

The amplicon produced from the F2/R2 primers permitted Sanger sequencing across aa 169 to aa 250 of the reverse transcriptase (RT) domain of the Pol gene (figure 1, highlighted in red), which is the region critical to identification of RAS. Of those samples for which an amplicon was obtained, 5/25 contained known polymerase RAS (table 2). A further 4 samples contained minor peaks on Sanger traces at known RAS sites (table 2).

MinION sequencing direct from DSS

Having obtained clinically relevant sequence data by Sanger sequencing from DSS, we investigated whether comparable data could be obtained using MinION sequencing in a small proof-of-principle study. Three high titre samples were selected from the Brazilian cohort (Br1, Br2, Br3), along with three samples from a separate cohort sourced from the Kurdistan region of Iraq. The samples sourced from Iraq all had a viral load >10^8 IU/mL. All samples were successfully amplified and Sanger sequenced using F1/R1 and F2/R2 primer pairs initially. An attempt was made to amplify whole HBV genomes from DSS using our

direct PCR protocol and a previously published primer set [25], however this was unsuccessful regardless of viral load.

Amplicons of 1,274 bp were successfully generated using primers F3 and R3. A total of 185,349 raw reads were obtained for the 6 samples, ranging from 23869 to 41392 per barcode. Following adapter trimming and further filtering of erroneously long and short reads these counts ranged from 7676 to 11832 per barcode (table 3). A summary of raw sequencing reads acquired over time, and average quality score per read over time is included in figure 2.

De novo consensus sequence building

Consensus sequences were assembled using Canu. Following initial assembly, the only errors observed were single base deletions within homopolymers when compared to Sanger sequences, ranging from 5 to 12 deletions (data not shown). Following a single round of Nanopolish processing all consensus sequences were identical to their Sanger counterparts (figure 3).

Detection of minor variants

Several putative minor variants were detected following alignment of MinION reads with the *de novo* consensus sequence. The majority of these were filtered out by LoFreq. A single variant was detected in sample Iq3, an A > G switch in 19% of reads coding for N155D within the spacer region of polymerase. Inspection of the Sanger trace confirmed the presence of this minor variant.

Discussion

Methods to both genotype and characterise therapeutic and diagnostic mutations can be technically, logistically or financially challenging in regions of the world with a high prevalence of chronic HBV infection. We developed two workflows for generating sequencing data from DSS using Sanger and MinION platforms, obtaining fully analysed sequences in less than 48 hours for both methods from receipt of samples. Deep sequencing data were successfully generated for all six of the samples tested and, following *in silico* processing, produced consensus sequences that were identical to the sequences generated by Sanger sequencing.

While both sequencing methods involve specific expenses, once established they are cost-effective, require no cold chain, and the PCR achieved a lower limit of detection of approximately 600 IU/mL for small amplicons. This level of sensitivity makes both workflows viable options for informing clinical decisions as the recommended limit set for initiating therapy is 2,000 IU/mL [6, 7]. However, while the direct PCR method used here worked well for partial gene amplicons at low viral loads, it was not successful when applied to previously published primer sets for amplifying the entire HBV genome [25] (data not shown), suggesting traditional extraction methods may be preferable for larger (>1.5kb) amplicons.

Several RAS detected in this cohort are clinically significant. The Y100C mutant in the S region of HBsAg has been linked to false negative/low HBsAg tests interpreted as occult infection [26]. This variant was unusually common in our Brazilian cohort, being present in 5/10 genotype A samples, compared to only 2.8% of genotype A samples analysed in a separate study, from South Africa [27].

Our PCR approach specifically targeted the RT domain of Pol to characterise drug resistance mutants. This is particularly important for application in countries where drugs with a low barrier to resistance are commonly prescribed. The V173L, L180M and M204V RAS detected in the polymerase gene indicated resistance to the nucleoside analogues lamivudine and entecavir [28]. The N236S mutation has not been previously described, but other mutations at this site have been shown to lead to breakthrough on adefovir [29] and decrease the efficacy of tenofovir *in vitro* [30]. Further research is required to determine if N236S produces a similar phenotype. Finally M250I, present in one sample as a heterozygous peak in Sanger sequencing, has been shown to lead to lamivudine and telbivudine resistance [28].

Complex selection pressures on HBV lead to evolution of intra-host subpopulations [31]. Accurately determining minor variants and haplotypes is potentially of great clinical value, particularly for a virus requiring long-term therapy allowing the expansion of multi-drug resistant haplotypes. By applying a conservative variant calling approach we identified one minor variant sequence using the MinION workflow (N155D in sample Iq3). This variant was subsequently confirmed by visual analysis of the Sanger read. As residue 155 is located in the spacer region of polymerase it is unlikely to be clinically significant but highlights the potential for detecting low-level variants. Multi-drug resistant populations can arise through sequential RAS introduction over time [32]. Long read high-throughput sequencing technologies such as nanopore allow characterisation of the changing population of HBV variants in high titre samples, which has previously required clonal analysis combined with Sanger sequencing.

We acknowledge a number of limitations of this study. Large amplicons were chosen to utilise nanopore sequencing directly from DSS in this proof-of-concept study, but greater sensitivity could be achieved using a PCR-tiling approach of two or more amplicons that are pooled before sequencing (as demonstrated with the much larger Zika virus genome [14]), or by scaling up reaction mixtures and using more DSS per reaction. Identical sequences were also observed in two pairs of high viral load samples chosen for nanopore sequencing. To confirm that this was not an artefact introduced by cross-contamination during processing, additional, independent DSS punches were taken and sequenced following direct PCR amplification from these samples to verify the sequences. Given the low substitution rate of HBV and the fact that these pairs of patients were sampled from the same geographic region it is likely that these are genuinely identical sequences. However, errors in preparing DSS cards cannot be discounted. Although we observed no carryover between high titre samples and those with an undetectable viral load, more stringent decontamination between punches could be considered if this workflow was applied in a clinical setting. We also highlight issues with obtaining amplicons >1.5kb using the direct PCR enzyme, which limits the ability of this direct method to detect linked mutations in long genomes.

In summary, we describe two approaches for rapid genotyping and RAS detection in HBV, using a novel analyte, dried serum spots, which are applicable in resource-limited settings and require little existing infrastructure. The results presented here demonstrate the utility of direct PCR enzymes and DSS together in a clinical context. We have also demonstrated, for the first time, that nanopore sequencing can be applied directly to samples amplified

from DSS, with no requirement for extraction. Reliable sequence data was generated using the MinION sequencer, significantly reducing the requirements for laboratory infrastructure.

Funding

Funding was provided by the Nottingham Molecular Pathology Node (MRC/EPSRC grant MR/N005953/1) and the National Institute for Health Research (NIHR) Nottingham Biomedical Research Centre.

Competing interests

The authors have no competing interests to declare.

References

[1] Lemoine M, Eholie S, Lacombe K. Reducing the neglected burden of viral hepatitis in Africa: strategies for a global approach. J Hepatol. 2015;62:469-76.

[2] Graber-Stiehl I. The silent epidemic killing more people than HIV, malaria or TB. Nature. 2018;564:24-6.

[3] O'Hara GA, McNaughton AL, Maponga T, Jooste P, Ocama P, Chilengi R, et al. Hepatitis B virus infection as a neglected tropical disease. PLoS Negl Trop Dis. 2017;11:e0005842.

[4] Lapinski TW, Pogorzelska J, Flisiak R. HBV mutations and their clinical significance. Adv Med Sci. 2012;57:18-22.

[5] Thompson AJ, Ayres A, Yuen L, Bartholomeusz A, Bowden DS, Iser DM, et al. Lamivudine resistance in patients with chronic hepatitis B: role of clinical and virological factors. J Gastroenterol Hepatol. 2007;22:1078-85.

[6] Liver EAftSoT. EASL 2017 Clinical Practice Guidelines on the management of hepatitis B virus infection. J Hepatol. 2017;67:370-98.

[7] Terrault NA, Lok ASF, McMahon BJ, Chang KM, Hwang JP, Jonas MM, et al. Update on prevention, diagnosis, and treatment of chronic hepatitis B: AASLD 2018 hepatitis B guidance. Hepatology. 2018;67:1560-99.

[8] Han GR, Cao MK, Zhao W, Jiang HX, Wang CM, Bai SF, et al. A prospective and open-label study for the efficacy and safety of telbivudine in pregnancy for the prevention of perinatal transmission of hepatitis B virus infection. J Hepatol. 2011;55:1215-21.

[9] Xu WM, Cui YT, Wang L, Yang H, Liang ZQ, Li XM, et al. Lamivudine in late pregnancy to prevent perinatal transmission of hepatitis B virus infection: a multicentre, randomized, double-blind, placebo-controlled study. J Viral Hepat. 2009;16:94-103.

[10] Barth RE, Huijgen Q, Taljaard J, Hoepelman AI. Hepatitis B/C and HIV in sub-Saharan Africa: an association between highly prevalent infectious diseases. A systematic review and meta-analysis. Int J Infect Dis. 2010;14:e1024-31.

[11] Matthews PC, Geretti AM, Goulder PJ, Klenerman P. Epidemiology and impact of HIV coinfection with hepatitis B and hepatitis C viruses in Sub-Saharan Africa. J Clin Virol. 2014;61:20-33.

[12] Weber B. Genetic variability of the S gene of hepatitis B virus: clinical and diagnostic impact. J Clin Virol. 2005;32:102-12.

[13] Quick J, Loman NJ, Duraffour S, Simpson JT, Severi E, Cowley L, et al. Real-time, portable genome sequencing for Ebola surveillance. Nature. 2016;530:228-32.

[14] Quick J, Grubaugh ND, Pullan ST, Claro IM, Smith AD, Gangavarapu K, et al. Multiplex PCR method for MinION and Illumina sequencing of Zika and other virus genomes directly from clinical samples. Nat Protoc. 2017;12:1261-76.

[15] Loman NJ, Quick J, Simpson JT. A complete bacterial genome assembled de novo using only nanopore sequencing data. Nat Methods. 2015;12:733-5.

[16] McNaughton AL, Roberts HE, Bonsall D, de Cesare M, Mokaya J, Lumley SF, et al. Illumina and Nanopore methods for whole genome sequencing of hepatitis B virus (HBV). Scientific Reports. 2019;9:7081.

[17] Kumar S, Stecher G, Li M, Knyaz C, Tamura K. MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. Mol Biol Evol. 2018;35:1547-9.

[18] Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 2004;32:1792-7.

[19] De Coster W, D'Hert S, Schultz DT, Cruts M, Van Broeckhoven C. NanoPack: visualizing and processing long-read sequencing data. Bioinformatics. 2018;34:2666-9.
[20] Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. Genome Res. 2017;27:722-36.
[21] Li H. Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics. 2018;34:3094-100.
[22] Beggel B, Neumann-Fraune M, Doring M, Lawyer G, Kaiser R, Verheyen J, et al. Genotyping hepatitis B virus dual infections using population-based sequence data. J Gen Virol. 2012;93:1899-907.
[23] Rhee SY, Margeridon-Thermet S, Nguyen MH, Liu TF, Kagan RM, Beggel B, et al. Hepatitis B virus reverse transcriptase sequence variant database for sequence analysis and mutation discovery. Antiviral Res. 2010;88:269-75.
[24] Wilm A, Aw PP, Bertrand D, Yeo GH, Ong SH, Wong CH, et al. LoFreq: a sequence-quality aware, ultra-sensitive variant caller for uncovering cell-population heterogeneity from high-throughput sequencing datasets. Nucleic Acids Res. 2012;40:11189-201.
[25] Gunther S, Li BC, Miska S, Kruger DH, Meisel H, Will H. 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. 1995;69:5437-44.
[26] Martin CM, Welge JA, Rouster SD, Shata MT, Sherman KE, Blackard JT. Mutations associated with occult hepatitis B virus infection result in decreased surface antigen expression in vitro. J Viral Hepat. 2012;19:716-23.
[27] Gencay M, Vermeulen M, Neofytos D, Westergaard G, Pabinger S, Kriegner A, et al. Substantial variation in the hepatitis B surface antigen (HBsAg) in hepatitis B virus (HBV)-positive patients from South Africa: Reliable detection of HBV by the Elecsys HBsAg II assay. J Clin Virol. 2018;101:38-43.
[28] Zoulim F, Locarnini S. Hepatitis B virus resistance to nucleos(t)ide analogues. Gastroenterology. 2009;137:1593-608 e1-2.
[29] Sheldon J, Camino N, Rodes B, Bartholomeusz A, Kuiper M, Tacke F, et al. Selection of hepatitis B virus polymerase mutations in HIV-coinfected patients treated with tenofovir. Antivir Ther. 2005;10:727-34.
[30] Brunelle MN, Jacquard AC, Pichoud C, Durantel D, Carrouee-Durantel S, Villeneuve JP, et al. Susceptibility to antivirals of a human HBV strain with mutations conferring resistance to both lamivudine and adefovir. Hepatology. 2005;41:1391-8.
[31] McNaughton AL, D'Arienzo V, Ansari MA, Lumley SF, Littlejohn M, Revill P, et al. Insights From Deep Sequencing of the HBV Genome-Unique, Tiny, and Misunderstood. Gastroenterology. 2019;156:384-99.
[32] Yim HJ, Hussain M, Liu Y, Wong SN, Fung SK, Lok AS. Evolution of multi-drug resistant hepatitis B virus during sequential therapy. Hepatology. 2006;44:703-12.

Figure captions

Figure 1. Cartoon highlighting the locations of PCR amplicons. Overlapping ORFs in the HBV genome (genotype A reference isolate X02763) are shown in blue. PCR amplicons generated in this study are shown in orange, and the crucial RT region associated with treatment resistance mutations (aa169 – 250) is shown in red. The F1+R1 amplicon is sufficient for genotyping and limited detection of sAg diagnostic escape mutants. The region of the reverse transcriptase (RT) domain in which resistance associated substitutions (RAS) arise (aa169 – 250), shown in red, is encompassed by the F2+R2 amplicon.

Figure 2. Metrics from MinION sequencing run. Both plots were generated from raw reads assigned barcodes by Albacore without further filtering. Quality scores are standard Phred scores produced by Albacore during basecalling, data is presented as mean Phred score per read with min and max.

Figure 3: Maximum likelihood tree comparing sequences sourced from direct PCR using both Sanger and nanopore sequencing methods. Nanopore data was first analysed using Nanopolish, before comparing to Sanger sequences. Sequences were generated using the amplicons generated from the F3/R3 primers. The ML tree was inferred using a general time reversible model within MEGA X [17]. Statistical robustness was assessed using bootstrap resampling of 1,000 pseudoreplicates. The tree with the highest log-likelihood is shown.

Table 1

Table 1: Primers used for the amplification of HBV. R2 is a modified version of the HBV3 primer. F3 and R3 primers contain additional 5' bases for MinION library preparation PCR. *Numbering based upon HBVdb genotype A reference strain X02763 [33].

Original name	This study	Sequence (5'-3')	HBV genome position*
Outer plus [34]	F1	GATGTGTCTGCGGCGTTTTA	376 - 395
Outer minus [34]	R1	CTGAGGCCCACTCCCATAGG	656-637
-	F2	GGAYGGAAAYTGCACYTGTA	583 – 602
-	R2	GRGCAACRGGGTAAAGG	1156 - 1140
HBVZ [35]	F3	TTTCTGTTGGTGCTGATATTGCAGCCCTCAGGCTCAGGGCATA	3085 – 3105
HBV3 [35]	R3	ACTTGCCTGTCGCTCTATCTTCCGTTGCCKDGCAACSGGGTAAAGG	1163 - 1140

Table 2: Amplification of S/Pol gene from DSS of HBV-positive samples from Brazil, allowing genotyping and characterisation of RAS by Sanger sequencing, is dependent on viral load. Clinically significant RAS in the reverse transcriptase (RT) domain of the Polymerase open reading frame (ORF) are noted. Treatment information: TDF, tenofovir; 3TC, lamivudine; EFV, efavirenz; U, treatment unknown. Amplification and sequencing information: Y, amplification and sequencing successful for given primer set; N, unsuccessful amplification for given primer set; NT, sample not tested due to lack of source material; ND, no RAS detected in given ORF. Genotype and % confidence data were obtained using HBV geno2pheno with F1+R1 sequence data. In the case of sample Br26, the consensus sequence was basecalled as R (A/G) at RT site 750, with G leading to the M250I mutant and A matching the genotype D reference sequence. The minor peaks column highlights any minority sequence changes at clinically significant RAS sites.

Sample	Viral load (IU/mL)	Therapy	F1+R1 sequence	Genotype	% confidence	F2+R2 sequence	RT mutants	sAg escape mutants	Minor peaks
Br1	816,483,772	No treatment	Y	A1	97.6	Y	ND	Y100C	ND
Br2	389,547,123	Peg-IFN	Y	A1	97.6	Y	ND	Y100C	ND
Br3	241,145,063	TDF, 3TC	Y	D2	98.9	Y	V173L, L180M M204V	ND	ND
Br4	33,875,365	TDF, 3TC, EFV	Y	A1	97.6	Y	ND	Y100C	ND
Br5	21,807,740	No treatment	Y	D3	98.5	Y	ND	ND	ND
Br6	6,219,005	TDF	Y	D3	98.7	Y	ND	ND	ND
Br7	5,843,310	No treatment	Y	A1	97.1	Y	ND	Y100C	ND
Br8	5,217,998	U	Y	D2	98.8	Y	ND	ND	ND
Br9	3,166,098	TDF	Y	F2	97.6	Y	V173L	ND	rtL173V, rtL180M, rtA181T, rtM204V
Br10	692,965	U	Y	A2	99.0	Y	ND	ND	ND
Br11	193,271	TDF	Y	F2	97.6	Y	ND	ND	ND
Br12	94,152	U	Y	D3	97.9	Y	N236S	ND	ND
Br13	59,927	TDF	Y	D3	98.6	N	ND	ND	ND
Br14	43,005	TDF, 3TC	Y	A1	97.1	Y	ND	Y100C	ND
Br15	35,672	TDF	Y	A1	97.6	Y	ND	ND	ND
Br16	25,791	TDF	Y	D2	95.1	Y	S202I	ND	ND
Br17	17,280	U	Y	F1	97.3	Y	ND	ND	ND
Br18	16,697	U	Y	D3	98.7	NT	ND	ND	ND
Br19	12,446	3TC	Y	D3	98.7	Y	ND	ND	rtI169M, rtM204L
Br20	12,196	No treatment	Y	A2	96.2	Y	ND	ND	ND
Br21	10,058	TDF	Y	F2	97.2	Y	ND	ND	rtI169K, rtV173G
Br22	7,199	TDF	N			N			
Br23	6,208	No treatment	Y	D3	98.7	Y	ND	ND	ND
Br24	4,091	TDF	Y	B1	98.5	Y	ND	ND	rtV173G
Br25	2,884	Peg-IFN	Y	A1	97.5	N			
Br26	2,431	No treatment	Y	D1	97.1	Y	M250I (A/G heterozygous)	ND	ND
Br27	2,262	No treatment	Y	A2	97.7	Y	ND	ND	rtN263K
Br28	1,069	U	N	A		Y	ND	ND	ND
Br29	1,046	No treatment	N			N			
Br30	635	No treatment	Y	D3	92.6	Y	ND	ND	ND

Table 3. MinION sequencing yields. Raw reads are those assigned barcodes by Albacore before any further quality control. Adapter trimmed reads are those exceeding a mean Phred score of 7 and processed by Porechop. Length filtered reads were processed by NanoFilt.

Sample	Viral load (IU/mL)	Raw reads	Adapter trimming	Length filtered
Br1	816,483,772	24105	13320	11272
Br2	389,547,123	23868	12640	10309
Br3	241,145,063	41392	14702	11832
Iq1	369,094,710	27478	12445	7676
Iq2	480,199,200	40646	13539	8979
Iq3	361,383,300	27860	11654	9572

Figure 1

Click here to download high resolution image

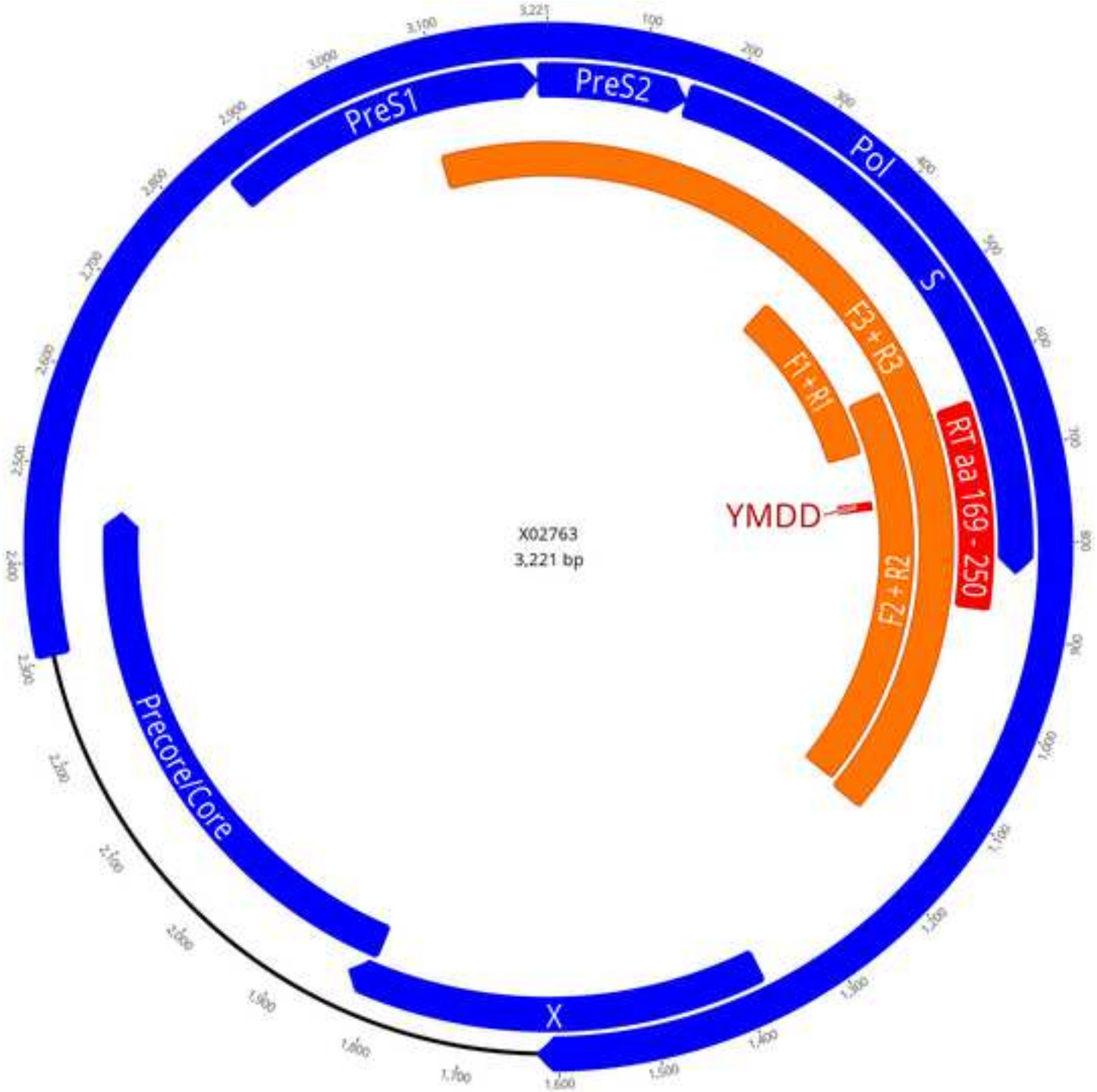
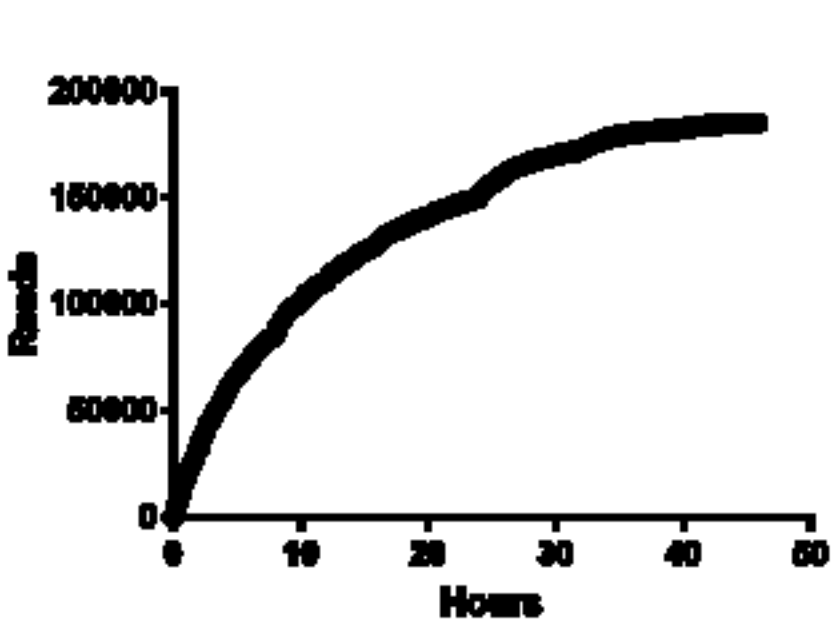


Figure 2

Click here to download high resolution image



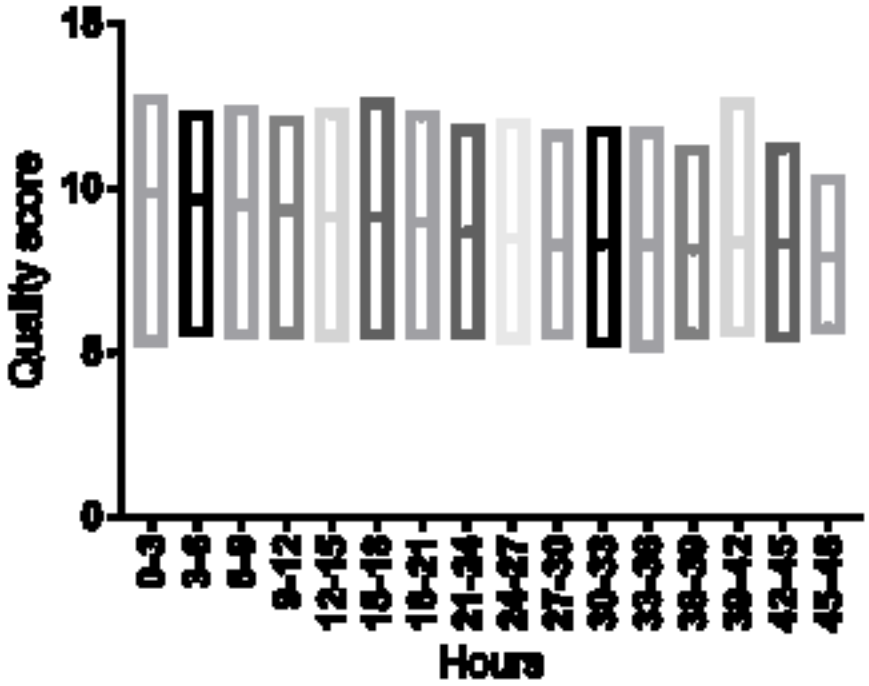
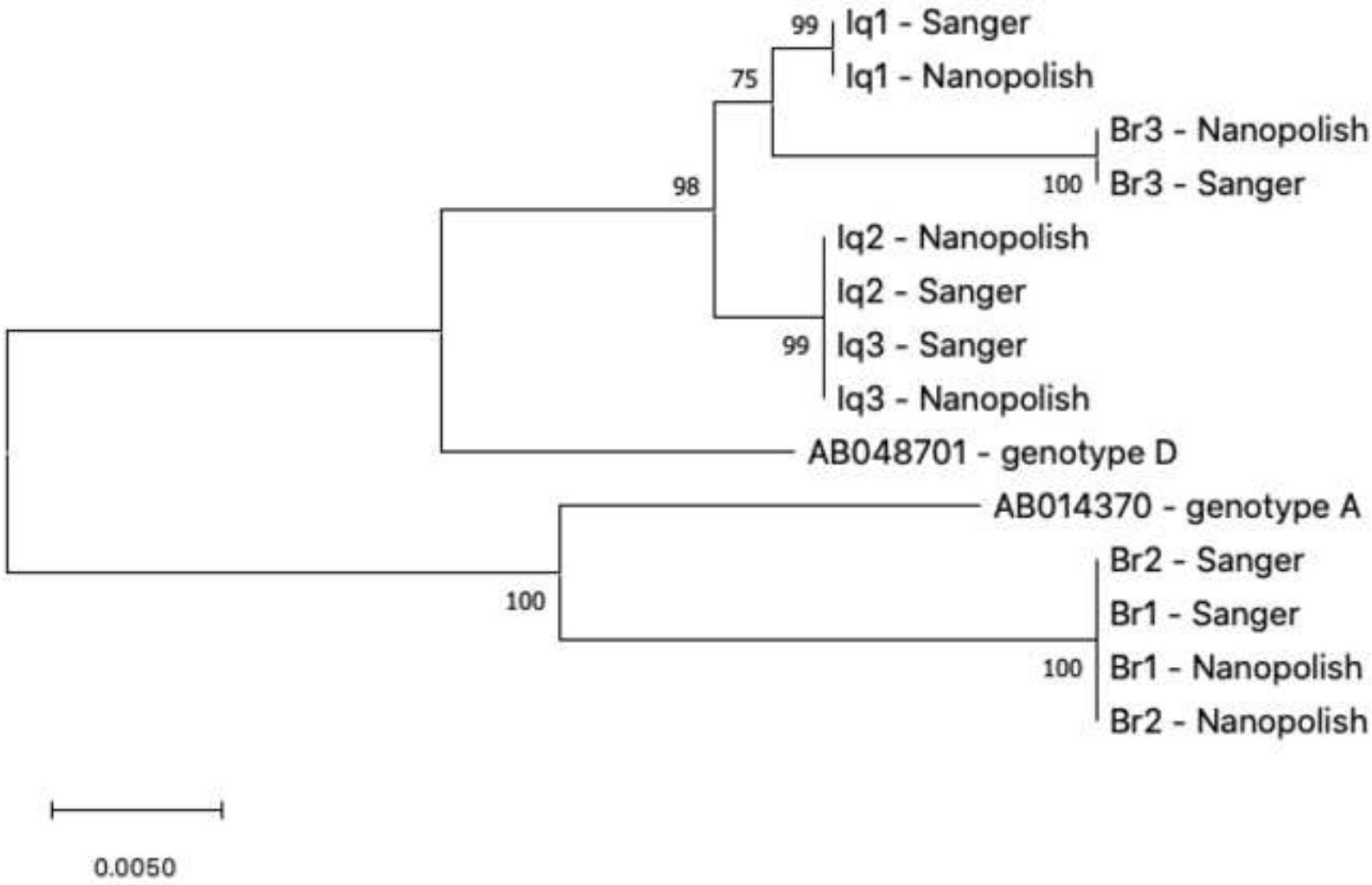


Figure 3 revised

Click here to download high resolution image



Supplementary data

Bioinformatics workflow

Following basecalling using Albacore, the “pass” folder (reads exceeding a mean Phred score of 7) was used as the input for Porechop as follows:

porechop -i source_directory --discard_middle --require_two_barcodes

Filtering based on length was carried out for each barcode using NanoFilt:

cat reads.fastq | nanofilt -l 1200 --maxlength 1300 > reads.filtered.fastq

Downsampled, filtered FASTQ files for each sample were then used as the input for Canu:

./canu --nanopore-raw reads.filtered.fastq genomeSize=1300 stopOnReadQuality=false -d canu_out -p sample-ID

This generates several candidate contigs, the contig with the highest number of reads used was verified using BLAST and taken forwards to the next step.

Reads were then aligned in Minimap2 to their respective contig generated using Canu:

minimap2 -ax map-ont canu_contig.fasta reads.filtered.fastq | samtools view -bS - | samtools sort -o sample.minimap.sorted.bam

This alignment was then used to generate a polished consensus sequence using Nanopolish. First the reads are indexed to match every read in the .fastq file with its corresponding raw fast5 file (the original output of the minION sequencer):

nanopolish index -d fast5_directory -s sequencing_summary.txt reads.filtered.fastq

The alignment, reference contig and fastq for each sample were then used as the input for Nanopolish:

nanopolish variants --consensus --fix-homopolymers -b sample.minimap.sorted.bam -g canu_contig.fasta -r reads.filtered.fastq -o sample.polished.consensus.vcf

This consensus .vcf file was then converted to standard .fasta format:

nanopolish vcf2fasta -g canu_contig.fasta sample.polished.consensus.vcf > sample.polished.consensus.fasta

The output consensus sequence can then be checked against the original Canu contig, as well as a Sanger contig from the same sample if available. These sequences can also be used for genotyping and resistance typing against established reference sequences.

Intra-sample variants can be determined by aligning all reads to the consensus sequence and using LoFreq to generate a .vcf file:

lofreq call-parallel --pp-threads 8 -f reference_genome.fas -o variants_file.vcf sample_alignment.bam

***Credit Author Statement**

Author contributions

CPM, AWT, WLI and SA designed the experiments; MMCNS, ACGJ, JFS, PJ, CHS and FTS provided samples and associated clinical data and prepared DSS cards; SA, EP and CPM carried out the experimental work; SA carried out nanopore sequencing and associated bioinformatics; SA, BK, AWT and CPM analysed the data and drafted the manuscript; all authors edited the manuscript.