Accurate and cost-efficient whole genome sequencing of hepatitis B virus using Nanopore

Joakim B. Stenbäck[1,2], Daniel Schmidt[3], Ulrika Noborg[1], Joel Gustafsson[1], Peter Norberg[1], Maria E. Andersson[1,2], Michael X Fu[4], Heli Harvala[5] and Johan Ringlander[1,2]

[1]Department of Infectious Diseases, Institute of Biomedicine, University of Gothenburg, Sweden

[2]Department of Clinical Microbiolgy, Sahlgrenska University Hospital, Gothenburg, Sweden

[3]Bioinformatics and Data Centre, University of Gothenburg, Sweden

[4]Nuffield Department of Medicine, University of Oxford, United Kingdom

[5]Radcliffe Department of Medicine, University of Oxford, United Kingdom

Abstract

Deep sequencing of the whole hepatitis B virus genome increases the analytical resolution and has the potential to improve molecular epidemiology investigations. The aim of this work was to develop and evaluate the performance of such deep sequencing using the Nanopore technology. The method includes an initial PCR step to generate two overlapping amplicons that cover the whole HBV genome, followed by sequencing using the Nanopore rapid barcoding kit that allows parallel analysis of several samples in one reaction. The libraries can be sequenced with the standard Nanopore flow cell on MiniIon or GridIon devices, as well as the Flongle. The performance of the method was evaluated by comparing Nanopore and Sanger sequences or qPCR results from 64 clinical samples. The Nanopore-derived consensus sequences were, on average, 99.9% similar to those from Sanger sequencing and the full HBV genome was determined in samples with HBV DNA levels of approximately 3 $\log_{10}$ IU/mL with MagNA pure 96 extraction and < 2 $\log_{10}$ IU/mL using a high-volume manual extraction protocol on a subset of samples from patients with very low viral load (1.62-3.74 IU/mL). A perfect agreement with Sanger/qPCR-derived genotype was seen. The cost of sequencing per genome using the Nanopore method is low, ranging of 6-37 euros. We conclude that whole-genome sequencing of HBV with Nanopore is well suited for genomic characterization, antiviral resistance mutation analysis and genotyping of HBV in a routine laboratory setting.

Background

Chronic infection with hepatitis B virus (HBV) causes a heavy health burden worldwide (1). Chronic hepatitis B poses a risk for long-term complications such as cirrhosis and cancer, a risk that can be reduced by treatment with nucleotide analogs (2). HBV has nine genotypes (A-I) with relatively distinct geographical distribution, differing from each other by at least 8% of the ≈ 3200 nucleotides that comprise the genome. In addition, there are a large number of subgenotypes, some of which require whole genome comparison for accurate classification. Whole genome sequencing can also support molecular epidemiology investigations, for example, when evaluating potential transmission between subjects infected with the same genotype (3, 4). Identification of genotypes can be achieved by simplified methods such as real-time PCR (5), but new sequencing techniques have become almost as easy and allow relatively fast genotyping at a relatively low cost. Genotyping has clinical applicability since infection with genotype C is a risk factor for liver complications, and infection with genotype A is more susceptible to interferon treatment (6).
Although antiviral resistance against tenofovir and entecavir is relatively rare, resistance analysis is quite often of interest in patients with poor therapeutic response. Entecavir resistance typically occurs after previous treatment with lamivudine (7, 8). Resistance to tenofovir is very rare, but mutations leading to substitution 236T in reverse transcriptase, in combination with other mutations, have been suggested to lead to antiviral resistance (9, 10).

Nanopore sequencing has been widely used during the COVID-19 pandemic (11-14), and this technique is also increasingly used in bacteriology for antimicrobial resistance genotyping (15), 16s and whole-genome sequencing (16, 17). Many clinical microbiological laboratories that previously did not perform sequencing are now in possession of one or multiple Nanopore instruments (18). Apart from whole-genome sequencing of SARS-CoV-2 (19), methods for many other microbes have been launched (20-22). The relatively low upfront cost for the device, low cost of sequencing and simple protocols are benefits that have been described. On the other hand, a higher error rate, the need for high computer calculation capacity and specialized bioinformatics may discourage the introduction of the Nanopore method in a routine laboratory. In this article, we describe the development of a user-friendly method for Nanopore sequencing of HBV and a coherent bioinformatical pipeline adapted for clinical use in routine laboratories.

Methods

Patients and samples

Patient samples from three different cohorts were part of this study. Cohort 1 consisted of 38 HBV positive sera from individual patients that had been analyzed with a Sanger sequencing-based resistance testing method as part of routine diagnostics at the Department of Clinical Microbiology, Sahlgrenska University Hospital, Gothenburg. One sample from this cohort was quantified with qPCR and serially diluted to assess the analytical sensitivity. Cohort 2 was made out of 19 sera that had previously been genotyped using qPCR (5) in routine diagnostics, also at Sahlgrenska University Hospital. Cohort 3 consisted of 7 plasma samples from blood donor screening that all had low viral load and previously had been quantified using ultrasensitive PCR and genotyped using a Sanger sequencing-based method in the United Kingdom, as previously described (23) . Details on the number of samples, genotype and viral load are found in Table 1.

Nucleic acid extraction and HBV DNA quantification

For cohorts 1 and 2, which had been previously analyzed in routine diagnostics at the Sahlgrenska University Hospital, Gothenburg, Sweden, total NA extraction from 200 uL of blood/plasma was performed using the MagNA pure 96 (Roche, Basel, Switzerland), according to the manufacturer's instructions. For the low viral load panel (cohort 3), large-volume extraction from 5 mL of plasma was made using a modified protocol for the Roche High Volume NA extraction kit, as previously described (23). Four samples in cohort 3 were analyzed using both extraction methods for comparison. For cohorts 1 and 2, HBV DNA quantification was performed using the Cobas HBV test on the 6800 platform (Roche) according to the manufacturer's instructions; for cohort 3, quantification was performed as previously described (23).

Serial dilutions of one sample with an original concentration of 5.34 $\log_{10}$ IU/mL were sequenced with Nanopore to further assess the sensitivity. The sample was diluted in H_20 to 1:10, 1:100, 1:1000 and 1:10000. Only the first sets of primers for the whole genome method was used and the nested PCR covering the RT region was not applied. The dilutions were sequenced on a flongle flow cell.

PCR and Nanopore sequencing

PCRs intended to generate two overlapping amplicons covering the whole genome were designed. The primers used for amplicon 1 were 56F (CCTGCTGGTGGCTCCAGT) and 1827R (GAAAAAGTTGCATGGTGCTGGT), and for amplicon 2 primers 1689F (ACCGACCTTGAGGCCTACTTCA) and 262R (CCACCACGAGTCTAGACTCT). Identical settings and reagents were used in the two amplifications. To the reaction, 5 µL of sample and 1 µL of each primer in 10 µM dilutions were added to 10 µL of master mix containing SuperScript III polymerase (ThermoFisher, Waltham, MA, USA) and 10 µL H_2O. A denaturation step at 95°C was included prior to 40 cycles of PCR, with annealing at 60°C for 10 s and elongation at 72°C for 2 min.

Capillary gel electrophoresis was performed on a TapeStation 4200 (using the DT5000 kit, Agilent, Santa Clara, CA, USA), and if no visible bands were seen, a nested PCR with inner primers 251F (GTGGTGGACTTCTCTCAATTTTC) and 1801R (CAGACCAATTTATGCCTACAGCCT) was performed with the same PCR settings as described in the first PCR. In this study, only the amplicon covering the reverse transcriptase (RT) region was nested.

Amplicons were barcoded, and libraries were prepared with the Rapid Barcoding Kit 96 SQK-RBK110.96, attaching barcodes and adapters randomly on the amplicons using transposase (both kits from Oxford Nanopore Technologies, Oxford, UK). Samples were pooled and sequenced with the MinIon sequencing device (Oxford Nanopore Technologies) and the R10.4.1 flowcell (FLO-MIN114) or an R10.4.1 flongle (FLO-FLG114) (Oxford Nanopore Technologies). Cohort 1 samples were sequenced five at a time on a flongle flowcell to mimic the planned use in routine diagnostics. Cohort 2 and 3 samples were sequenced on one standard flowcell each, and reruns for reproducibility testing of cohort 3 samples were performed using the flongle.

Sanger sequencing

Parts of the reverse transcriptase (RT) region was sequenced with the Sanger method. For cohort 1, Sanger sequencing was performed after PCR with primers 593F(TGCACCTGTATTCCCATCCCATC) and 1010R(GRGCAGCAAADCCCAAAAGACC), if no band was seen, semi-nested PCR was performed with 635F (TTCCTATGGGAGTGGGCCTC) and the same reverse. For cohort 3, parts of the RT region was covered by two amplicons and Sanger sequenced, as previously described (23).

Bioinformatics

Base calling of Nanopore fast5 files was made with the GPU-accelerated guppy program (https://github.com/asadprodhan/GPU-accelerated-guppy-basecalling) with super accurate mode settings. Subsequently, fastq files were run through a pipeline including following steps: Reads were mapped to HBV reference genomes representing genotypes A-H using Minimap2. Draft consensus sequences were created using samtools and polished with Medaka (Oxford Nanopore Technologies) and the consensus sequence from the best matching genotype was chosen. Minor variant analysis was performed with freebayes (24) and a an inhouse script was used to call resistance mutations based on those listed at geno2pheno (https://hbv.geno2pheno.org/). The pipeline is available at github (https://github.com/ClinicalGenomicsGBG/hbv_nano). The Stanford HBV database (https://hivdb.stanford.edu/HBV/HBVseq/development/hbvseq.pl?action=showSequenceForm) and geno2pheno tool (https://hbv.geno2pheno.org/) were used for genotyping and calling antiviral resistance mutations from consensus sequences and to compare sequence quality in the RT region, and showed consistent results with the pipeline. Consensus sequences from Nanopore and Sanger were aligned with the alignment tool in CLC Genomics Workbench (Qiagen, Aarhus, Denmark) using standard settings and phylogenetic trees were created using the Jukes-Cantor method with 1000 bootstraps. Comparison of sequences was performed with the sequence comparison tool of CLC Genomics Workbench. Splicing analysis was made using FLAME (25).

Results

Antiviral resistance analysis and genotyping

In cohort 1, 18/38 samples had previously detected antiviral resistance mutations by Sanger sequencing. The same mutations were found using Nanopore, with the addition of a mutation causing the resistance-associated amino acid substitution 80I in one sample, that only was found with the Nanopore method, as this position was not covered by the Sanger amplicon. Samples with resistance mutations had a viral load of 5.0 $\log_{10}$ IU/mL in mean (range 3-7.92 $\log_{10}$ IU/mL) All minor variants found with the Sanger-based analysis were also detected by Nanopore and the antiviral resistance mutation with the lowest frequency was present in 25% of the Nanopore reads. All samples but four could be analyzed regarding antiviral resistance without performing nested PCR and those had low viral loads ranging 2.4-3.37 $\log_{10}$ IU/mL. Thus, the clinical sensitivity for antiviral resistance mutations was 100%. There were no

antiviral resistance-associated mutations in cohorts 2 ands 3, that had previously been analyzed for resistance mutations.

In cohort 1,2 and 3, genotyping using Sanger or qPCR (cohort 2) did not differ from the Nanopore results. HBV genotypes A-F were represented among the sample (table 1). The number of samples with mismatches in the Nanopore method primer targets was low and typically only in one position. However, the four samples with an unsuccessful whole genome sequencing might have had more mismatches in primers 56F and 1827R, that could not be detected. Sequencing of the RT region of these samples after nested PCR was successful and showed a lack of mutations in 1689F and 262R primer targets for all but the one sample with genotype F that had multiple differences.

Sensitivity and reproducibility

In cohort 1, the entire HBV genome was obtained in all samples >3.27 log10 IU/mL, that was 33/38 samples. In the one sample with genotype F, the whole amplicon covering the RT region, but not the whole genome, was obtained after the first PCR, thus, nested PCR was not performed. The remaining four samples (all ≤3.27 log10 IU/mL with genotype D (n=3) and genotype B (n=1)) were subject to nested PCR targeting the region spanning nt 251-1801(including the RT), and the entire region was obtained in all samples.

A sample with an original concentration of 5.34 $\log_{10}$ IU/mL was diluted and sequenced. Sequences were identical in dilutions down to 1:100 but the number of reads in 1:1000 and 1:10000 were insufficient for antiviral resistance analysis and assembly of a whole genome. At dilution 1:100, the full genome was not captured but the whole RT region was, which was identical to the RT region of the 1:1 and 1:10 dilutions. Thus, the analytical sensitivity of the method, using the standard MagNA pure 96 extraction and when five samples were pooled and run on one flongle flow cell, was approximately 3.34 $\log_{10}$ IU/mL. Table 2 shows results from the sequencing of a serially diluted sample.

In cohort 2 all samples were successfully whole genome sequenced.

For cohort 3, consisting of seven samples, of which four samples were extracted with both the automated MagNA pure 96 protocol and the large-volume manual extraction protocol, the full HBV genome was successfully Nanopore sequenced after manual extraction in all samples. The full genome was captured in 3/4 of the samples after MagNA pure 96 extraction, and in the fourth sample, only the amplicon covering the RT region showed reads. The three successful samples had viral loads of 3.43, 3.27 and 1.74 $\log_{10}$ IU/mL, respectively, and the

remaining sample had a viral load of 1.62 log IU/mL. Two of the four high volume extracted samples in cohort 3 (with viral load 3.37 and 1.62 $\log_{10}$ IU/mL) were run on both the flongle R10.4 flowcell (FLO-FLG114) and the regular R10.4 flowcell, (FLO-MINI 114) to investigate reproducibility, and consensus sequences were identical from both flowcells.

Sequence similarity

In cohort 1, the sequence similarity between Nanopore and Sanger was 99.9% (range 99.8-100%). Cohort 2 had not been Sanger sequenced. In cohort 3, the Sanger sequences covering the variable part of the RT were compared with the same regions in the Nanopore-derived sequences and showed a 99.8% similarity in mean (99.13-100%).

HBV splice variants and core deletions

Splicing at conventional splice sites was not detected in frequencies ≥10% threshold in any of the samples, however, reads suggesting splicing was found in 29 samples when the frequency threshold was lowered to 1%. Some of the HBV splice sites may affect the PCR target, most importantly the nt 2447-489 splicing junction, and could therefore potentially not be detected. Detection of all known preS and preCore deletions (26, 27) is however possible. Notably, in cohort 3, consisting of low viral load samples, one sample showed a preS1 deletion spaning nt 2847-3060, as the predominant genome variant, and one sample harboring an nt 2994-3054 deletion as a minor variant in approximately 25% of the reads (figure 2). No deletions were seen in cohorts 1 and 2.

Nanopore sequencing cost

Nanopore sequencing with the method described in this article can be performed with up to 96 samples per flow cell run or approximately twelve samples using the flongle flow cell. With current costs for conventional extraction and PCR (4€) and Nanopore reagents (R10.4.1 flow cell from 455€, flongle flow cell from 66€ and rapid barcoding kit cost per total barcoded library of 99€) the reagent cost for each run is in the range of 169-558 €, thus for each genome in the range of 6-37€, depending on flow cell used and the number of samples pooled.

Discussion

Nanopore whole-genome sequencing of HBV was fast, sensitive and accurate, with 99.9% sequence agreement with Sanger sequencing. By introducing barcodes with the nanopore rapid barcoding method, antiviral resistance, genotype and sequence analyses can be achieved at low cost. The main advantages of Nanopore are its simplicity and its feature to perform long-read sequencing. By amplifying and sequencing only two amplicons with lengths of approximately 1771 nt (nt 56 to 1827) and 1773 nt (nt 1689 to 262), respectively, the whole circular HBV genome was captured. The primers of the two PCRs targeted conserved regions, thus, all major genotypes could be sequenced with the same set of primers and reagents. The primer regions were chosen to obtain continuous reads for the RT, precore-core region and for the regions potentially involved in vaccine escape and antiviral resistance (28-30). For this study, we used the Rapid Barcoding kit 96 V14 but alternative Nanopore kits are also possible. The ligation sequencing-based kits were tested during the method development but tend to be more laborious and time consuming, however, the amplicons remain intact and are sequenced as one continuous read, which may be important for haplotype and structural variation analysis. The primers used for this method can also easily be adapted for Nanopore rapid barcoding PCR kits, enabling fusion PCR of adapter and barcode sequences and thereby a simplified workflow without having the transposase of the rapid barcoding kit shortening the reads. Despite this transposase effect, we believe that the rapid barcoding kit is best suited for routine diagnostics, as it demands little hands-on time and may be used for sequencing and pooling of several different microbes simultaneously, which may be either PCR-enriched or cultured. This is also the most used kit for SARS-CoV-2 sequencing and has proven to be scalable to up to 96 samples in one run.

We tested two different nucleic acid extraction protocols in this study, and the modified protocol for the manual high-volume kit from 5 mL of plasma by Fu *et al.* (23) proved to be useful in low viral load samples, and full genome coverage was achieved from samples as low as 1.62 log10 IU/mL, that could not be achieved with MagNA pure 96 extraction from microlitres of serum. Serial dilution of one sample showed that for concentrations >3.43 log10 IU/mL, the MagNA pure 96 (Roche) showed reproducible results and was in all samples of cohort 1 sufficient for resistance analysis and sequencing of the full RT gene. Further, the full genome was captured in eight samples with viral load <3.43 log10 IU/mL also after MagNA pure 96 extraction. However, we conclude that the manual kit was more

sensitive, presumably because it yielded higher levels of intact DNA and, therefore, was preferable for long-read sequencing.

Bioinformatical optimization is a crucial part of developing Nanopore based methods. Most importantly, we applied the GPU-based super accurate basecalling for fast yet high-quality generation of reads. Nanopore basecalling has improved immensely in recent years and is the main factor for lowered error rate and several alternative methods have been presented (31-34). In a previous study, we used a pipeline for consensus sequence extraction from NGS data and almost perfect sequences were achieved, but some false insertion/deletions (indels) remained (3). To overcome this issue in more error prone long-reads data, a pipeline based on the Oxford Nanopore Technologies tool Medaka creating an optimal consensus sequence from Minmap2-mapped reads was created. The implementation of the Medaka tool removed false indels, as previously shown for SARS-CoV-2 (19). Medaka creates consensus sequences based on the major virus variant and for minor variant analysis, we applied freebayes (24). In our analysis of antiviral resistance-related minor variants freebayes found all variants that previously had been found using Sanger sequencing and no false positives were seen, applying a frequency threshold of 10%. Our clinical samples had no resistance mutations with a frequency of <25%, which might support the threshold used, however, based on previous studies of error-rate, the 10% threshold could potentially be lowered to 3% (35), something we aim to investigate in future studies.

Previous studies have presented methods for sequencing HBV using rolling circle PCR (36, 37). McNaughton *et al.* adapted the method for Nanopore sequencing and minor variant analysis (37). Their approach with rolling-circle amplification resulted in extremely long amplicons and very long reads with the whole HBV sequence repeated. For routine diagnostics we found it more suitable to use a standard PCR approach rather than a rolling circle amplification that presumably has lower sensitivity (36) and more laborious protocols (37). Detection of very rare minor variants has not been shown to be clinically relevant and our method, based on conventional PCR, had an almost 100% agreement between Nanopore and Sanger derived consensus genomes, allowing for clinical use. For research purposes, if detection of ultra-rare minor variants is needed, we believe that the method presented by McNaughton *et al*, Ion Torrent or Illumina-based methods, might be more suitable (38).

An advantage of long-reads sequencing is the detection of structural sequence differences in DNA. Our results show that some splicing variants and most known deletions in HBV genomes can be found using long-read sequencing. The existence of such aberrations has been known for many years but their clinical importance is relatively unknown (39). A recent article suggests that splicing is associated with poorer interferon treatment outcomes (40). HBV genome deletions, such as those we found in the S region in two samples, have been associated with an increased risk of cirrhosis (41). The method presented in this article can be suitable for both large-scale use and characterization of structural genomic features, which allows for bigger studies. Other primers than those presented here could improve the detection of splicing variants (41-43) and we plan to study this.
Other limitations of this study include the lack of Sanger sequencing covering the entire HBV genome, as we did not have the resources for this, meaning that other regions than the RT could not be compared. Also, the primers for the second amplicon in the Nanopore method, covering the S region, are presumably not optimal for genotypes F-J, but primer targets for the amplicon covering the RT appear to be conserved in all genotypes. The aim of this article was to present a method suitable for a high-throughput routine laboratory and thus only presents a nested PCR for the amplicon covering the RT region, however, the whole genome sensitivity may be further increased for all genotypes if additional primers or nesting of the second amplicon is applied.

Our study shows that Nanopore sequencing of two overlapping amplicons is a feasible method for HBV whole-genome sequencing, genotyping and antiviral resistance mutation analysis in a routine laboratory setting, showing adequate sensitivity and accuracy for clinical purposes. Minor variant analysis of clinically relevant subpopulations was possible. We believe that this method, providing simple and cost-efficient sequencing, may contribute to increased whole-genome sequencing of HBV and new discoveries of clinically relevant genomic features.

Author contributions

JR, HH and UN planned and designed the study. JR, MA and JBS designed the Nanopore method. JR and PN financed the study. JBS, UN, MA and MF performed laboratory work. JG designed hardware and set up GPU-based basecalling. DS, JR and JBS designed and performed bioinformatical analyses. JR drafted the manuscript. All authors contributed to the final version of the manuscript.

Funding

This study was funded by the Gothenburg Society of Medicine and the Sahlgrenska University Hospital.

References

1. Cao G, Liu J, Liu M. Trends in mortality of liver disease due to hepatitis B in China from 1990 to 2019: findings from the Global Burden of Disease Study. Chin Med J (Engl). 2022;135(17):2049-55.
2. Liaw YF. Reduction of cirrhosis and hepatocellular carcinoma with antiviral therapy in chronic hepatitis B. Hepatology. 2013;58(5):1856.
3. Ringlander J, Andersson ME, Prakash K, Larsson SB, Lindh M. Deep sequencing of hepatitis B virus using Ion Torrent fusion primer method. J Virol Methods. 2022;299:114315.
4. Eilard A, Andersson M, Ringlander J, Wejstal R, Norkrans G, Lindh M. Vertically acquired occult hepatitis B virus infection may become overt after several years. J Infect. 2019;78(3):226-31.
5. Malmstrom S, Berglin-Enquist I, Lindh M. Novel method for genotyping hepatitis B virus on the basis of TaqMan real-time PCR. J Clin Microbiol. 2010;48(4):1105-11.
6. Malmstrom S, Eilard A, Larsson SB, Hannoun C, Norkrans G, Lindh M. Genotype impact on long-term virological outcome of chronic hepatitis B virus infection. J Clin Virol. 2012;54(4):321-6.
7. Cassino L, Benetti S, Fay F, Tanno H, Quarleri J. Unsuccessful therapy with adefovir and entecavir-tenofovir in a patient with chronic hepatitis B infection with previous resistance to lamivudine: a fourteen-year evolution of hepatitis B virus mutations. BMC Infect Dis. 2011;11:178.
8. Singal AK, Fontana RJ. Meta-analysis: oral anti-viral agents in adults with decompensated hepatitis B virus cirrhosis. Aliment Pharmacol Ther. 2012;35(6):674-89.
9. Mokaya J, McNaughton AL, Bester PA, Goedhals D, Barnes E, Marsden BD, 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. 2020;5:151.
10. Park ES, Lee AR, Kim DH, Lee JH, Yoo JJ, Ahn SH, et al. Identification of a quadruple mutation that confers tenofovir resistance in chronic hepatitis B patients. J Hepatol. 2019;70(6):1093-102.
11. Davis JJ, Long SW, Christensen PA, Olsen RJ, Olson R, Shukla M, et al. Analysis of the ARTIC Version 3 and Version 4 SARS-CoV-2 Primers and Their Impact on the Detection of the G142D Amino Acid Substitution in the Spike Protein. Microbiol Spectr. 2021;9(3):e0180321.
12. Luo J, Meng Z, Xu X, Wang L, Zhao K, Zhu X, et al. Systematic benchmarking of nanopore Q20+ kit in SARS-CoV-2 whole genome sequencing. Front Microbiol. 2022;13:973367.
13. Mannsverk S, Bergholm J, Palanisamy N, Ellstrom P, Kaden R, Lindh J, et al. SARS-CoV-2 variants of concern and spike protein mutational dynamics in a Swedish cohort during 2021, studied by Nanopore sequencing. Virol J. 2022;19(1):164.

14. Bull RA, Adikari TN, Ferguson JM, Hammond JM, Stevanovski I, Beukers AG, et al. Analytical validity of nanopore sequencing for rapid SARS-CoV-2 genome analysis. Nat Commun. 2020;11(1):6272.
15. Taxt AM, Avershina E, Frye SA, Naseer U, Ahmad R. Rapid identification of pathogens, antibiotic resistance genes and plasmids in blood cultures by nanopore sequencing. Sci Rep. 2020;10(1):7622.
16. Widen J, Moren J, Molling P, Fagerstrom A, Enblad P, Eriksson BM, et al. Diagnosis of external ventricular drainage related infections with real-time 16S PCR and third-generation 16S sequencing. Infect Dis (Lond). 2024;56(7):521-30.
17. Salva-Serra F, Jaen-Luchoro D, Jakobsson HE, Gonzales-Siles L, Karlsson R, Busquets A, et al. Complete genome sequences of Streptococcus pyogenes type strain reveal 100%-match between PacBio-solo and Illumina-Oxford Nanopore hybrid assemblies. Sci Rep. 2020;10(1):11656.
18. Schmidt J, Blessing F, Fimpler L, Wenzel F. Nanopore Sequencing in a Clinical Routine Laboratory: Challenges and Opportunities. Clin Lab. 2020;66(6).
19. Arevalo MT, Karavis MA, Katoski SE, Harris JV, Hill JM, Deshpande SV, et al. A Rapid, Whole Genome Sequencing Assay for Detection and Characterization of Novel Coronavirus (SARS-CoV-2) Clinical Specimens Using Nanopore Sequencing. Front Microbiol. 2022;13:910955.
20. Zhang Y, Lu X, Tang LV, Xia L, Hu Y. Nanopore-Targeted Sequencing Improves the Diagnosis and Treatment of Patients with Serious Infections. mBio. 2023:e0305522.
21. Chan WS, Au CH, Leung SM, Ho DN, Wong EYL, To MY, et al. Potential utility of targeted Nanopore sequencing for improving etiologic diagnosis of bacterial and fungal respiratory infection. Diagn Pathol. 2020;15(1):41.
22. Chen Y, Mao L, Lai D, Xu W, Zhang Y, Wu S, et al. Improved targeting of the 16S rDNA nanopore sequencing method enables rapid pathogen identification in bacterial pneumonia in children. Front Cell Infect Microbiol. 2022;12:1001607.
23. Fu MX, Simmonds P, Andreani J, Baklan H, Webster M, Asadi R, et al. Ultrasensitive PCR system for HBV DNA detection: Risk stratification for occult hepatitis B virus infection in English blood donors. J Med Virol. 2023;95(10):e29144.
24. Erik Garrison GM. Haplotype-based variant detection from short-read sequencing. arXiv. 2012(1207.3907).
25. Holmqvist I, Backerholm A, Tian Y, Xie G, Thorell K, Tang KW. FLAME: long-read bioinformatics tool for comprehensive spliceome characterization. RNA. 2021;27(10):1127-39.
26. Gerken G, Kremsdorf D, Capel F, Petit MA, Dauguet C, Manns MP, et al. Hepatitis B defective virus with rearrangements in the preS gene during chronic HBV infection. Virology. 1991;183(2):555-65.
27. Marinos G, Torre F, Gunther S, Thomas MG, Will H, Williams R, et al. Hepatitis B virus variants with core gene deletions in the evolution of chronic hepatitis B infection. Gastroenterology. 1996;111(1):183-92.
28. Wilson JN, Nokes DJ, Carman WF. Current status of HBV vaccine escape variants--a mathematical model of their epidemiology. J Viral Hepat. 1998;5 Suppl 2:25-30.
29. Huang X, Lu D, Ji G, Sun Y, Ma L, Chen Z, et al. Hepatitis B virus (HBV) vaccine-induced escape mutants of HBV S gene among children from Qidong area, China. Virus Res. 2004;99(1):63-8.
30. Basuni AA, Carman WF. HBV vaccine-escape variants. Methods Mol Med. 2004;95:115-24.

31. Yeh YM, Lu YC. MSRCall: a multi-scale deep neural network to basecall Oxford Nanopore sequences. Bioinformatics. 2022;38(16):3877-84.
32. Wick RR, Judd LM, Holt KE. Performance of neural network basecalling tools for Oxford Nanopore sequencing. Genome Biol. 2019;20(1):129.
33. Neumann D, Reddy ASN, Ben-Hur A. RODAN: a fully convolutional architecture for basecalling nanopore RNA sequencing data. BMC Bioinformatics. 2022;23(1):142.
34. Napieralski A, Nowak R. Basecalling Using Joint Raw and Event Nanopore Data Sequence-to-Sequence Processing. Sensors (Basel). 2022;22(6).
35. Grubaugh ND, Gangavarapu K, Quick J, Matteson NL, De Jesus JG, Main BJ, et al. An amplicon-based sequencing framework for accurately measuring intrahost virus diversity using PrimalSeq and iVar. Genome Biol. 2019;20(1):8.
36. Martel N, Gomes SA, Chemin I, Trepo C, Kay A. Improved rolling circle amplification (RCA) of hepatitis B virus (HBV) relaxed-circular serum DNA (RC-DNA). J Virol Methods. 2013;193(2):653-9.
37. 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). Sci Rep. 2019;9(1):7081.
38. Ringlander J, Fingal J, Kann H, Prakash K, Rydell G, Andersson M, et al. Impact of ADAR-induced editing of minor viral RNA populations on replication and transmission of SARS-CoV-2. Proc Natl Acad Sci U S A. 2022;119(6).
39. Sommer G, Gunther S, Sterneck M, Otto S, Will H. A new class of defective hepatitis B virus genomes with an internal poly(dA) sequence. Virology. 1997;239(2):402-12.
40. Chen J, Wu M, Wang F, Zhang W, Wang W, Zhang X, et al. Hepatitis B virus spliced variants are associated with an impaired response to interferon therapy. Sci Rep. 2015;5:16459.
41. Gunther S, Sommer G, Iwanska A, Will H. Heterogeneity and common features of defective hepatitis B virus genomes derived from spliced pregenomic RNA. Virology. 1997;238(2):363-71.
42. Maslac O, Wagner J, Sozzi V, Mason H, Svarovskaia J, Tan S, et al. Secreted hepatitis B virus splice variants differ by HBV genotype and across phases of chronic hepatitis B infection. J Viral Hepat. 2022;29(8):604-15.
43. Sommer G, van Bommel F, Will H. Genotype-specific synthesis and secretion of spliced hepatitis B virus genomes in hepatoma cells. Virology. 2000;271(2):371-81.

Figures and legends

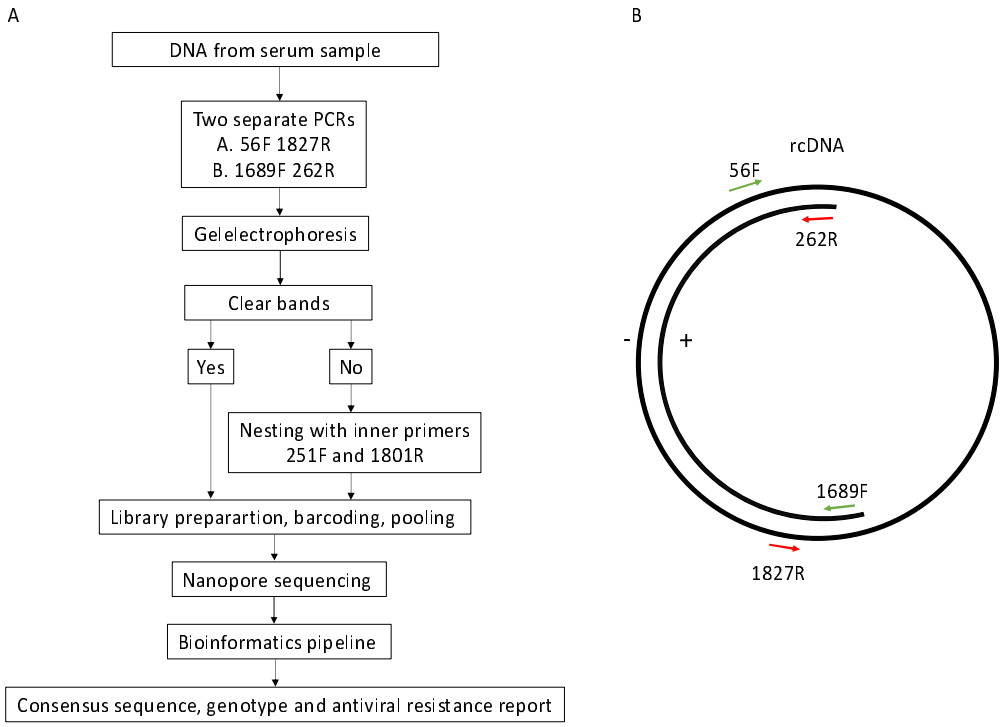

Figure 1. Method overview. A) Nanopore HBV whole-genome sequencing workflow. B) The position of the primers on the partially double-stranded HBV relaxed circular DNA (rcDNA), found in HBV particles. The method is designed to amplify both strands, including the shorter plus strand.

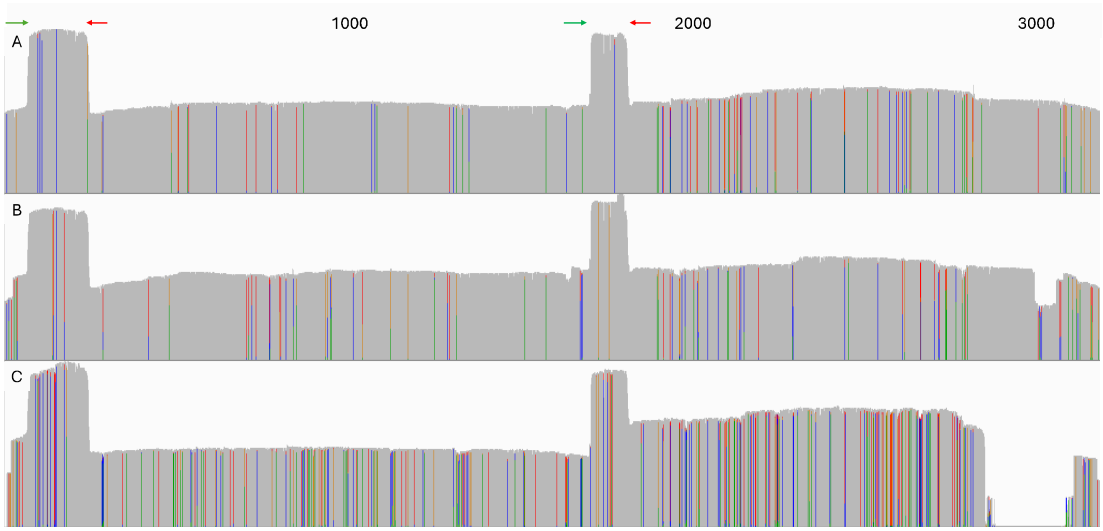

Figure 2. HBV coverage profiles from Nanopore sequencing of samples with and withour structural aberrations. Colored lines indicate differences compared to the reference sequence reads were mapped

against and grey matching nucleotides, the height reflects the number of reads covering the positions. The approximate primer locations are marked as green (forward) and red (reverse) arrows. A) Depicts a typical coverage profile, B) shows a sample with a preS1 deletion as a minor variant seen as a gap in a fraction of the reads at nt 2994-3054 and C) a sample with a majority of the HBV genomes with a preS1 deletion nt 2847-3060.

Tables

Table 1. Number of patients, genotype and viral load among the study cohorts.

HBV genotype		**Cohort 1**	**Cohort 2**	**Cohort 3**	**Total**
	A	5	0	2	7
	B	2	6	1	9
	C	8	8	1	17
	D	19	4	1	24
	E	3	1	2	6
	F	1	0	0	1
	All	38	19	7	64
Viral load IU/mL log_{10}					
	<3	2	0	4	5
	3-5	24	0	3	28
	>5	12	19	0	31
Same genotype as Sanger/qPCR		100%	100%	100%	100%
Antiviral resistance sensitivity		100%	NA	NA	100%
Viral load IU/mL log_{10}, mean		4.5	7.5	2.5	4.7
Fraction of the HBV genome, mean		99.9%	100%	99.8%	99.9%

Table 2. Analytical sensitivity and reproducibility after MagNA pure total NA extraction when running five pooled dilutions of one sample on a flongle flow cell.

Dilution	**log_{10} IU/mL**	**% of full genome**	**Similiarity**	**Antiviral resistance detected***
1:1	5,34	100	100	L180M, S202G, M204V
1:10	4,34	100	100	L180M, S202G, M204V
1:100	3,34	56	100	L180M, S202G, M204V
1:1000	2,34	0	NA	No
1:10000	1,34	0	NA	No

*Previous Sanger sequencing showed L180M, S202G and M204V.