Faculty Day 2014

Poster number: 2014234

FULL GENOME AMPLIFICATION AND QUASISPECIES RECONSTRUCTION OF THE HEPATITIS B VIRUS FROM NEXT GENERATION SEQUENCING DATA

Louis S. Le Clercq[1], Sheila M. Bowyer[1,2], Sim H. Mayaphi[1,2]
(1) Department of Medical Virology, University of Pretoria
(2) Tshwane Academic Division, National Health Laboratory Service

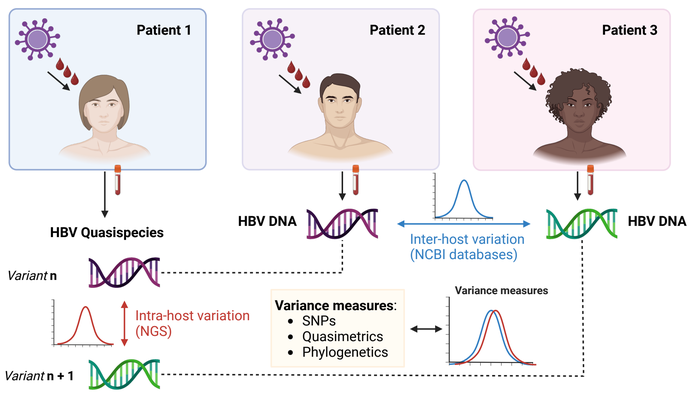
INTRODUCTION

The Hepatitis B Virus (HBV) is a DNA virus that infects most vertebrates and is part of the Hepadnaviridae family. Transmission of HBV occurs through bodily fluids such as blood and sexual contact. HBV targets the liver cells of mammals, including humans, and can result in acute or chronic infections. The virus replicates within host liver cells by reverse transcription of an RNA intermediate and also integrates into the host genome. HBV undergoes high mutation rates due to its error-prone replication, leading to a diverse population of viral variants. Consequently, It has several genotypes (A to J). In chronic infections, HBV forms a complex mixture of viral variants within the host, known as a quasispecies consisting of major, intermediate, and minor variants.

As selection occurs on the entire viral population, and progresses over the course of infection, the dynamics of viral evolution cannot be understood from the fittest strain alone. Therefor, quasispecies reconstructions present a unique avenue to gain novel insights into the emergence of novel strains. This research seeks to enhance our understanding of HBV quasispecies dynamics.

METHODS

- DNA was extracted from plasma samples on the MagNA Pure LC instrument or with the QIAamp MinElute Virus Spin kit
- Amplification for full-length HBV genome was done as described by Günther *et al.* (1995)
- Success of amplification was monitored by TBE-Agarose electrophoresis
- PCR amplicons were purified with the DNA Clean & Concentrator™-25 kit
- Purified amplicons and relevant controls were sequenced by next generation sequencing on the MiSeq sequencer (Illumina)



- The FASTQ sequence files were analysed for quality and assembled to a reference in GALAXY
- The viral quasispecies was reconstructed for each specimen with a validated Java executable algorithm QuRe V.0.99971
- Phylogenetic analysis was done with the Neighbour-Joining method in MEGA
- Variant calling of SNPs were done in GALAXY and VarScan for quasispecies

CONCLUSION

NGS and quasispecies reconstruction greatly enhances the ability to model viral evolution and has the potential to possibly predict future isolates.

RESULTS

- 30 samples were successfully amplified using the optimised PCR and primers from this study
- 15 samples were successfully sequenced using NGS
- Most samples had full genome coverage at a depth of more than 10X
- Read size ranged between 150 and 250 base pairs
- Per base sequence content showed variation in the first 15 base pairs but normalised to equal distributions
- The majority of reads passed the lower quality score threshold of 20
- Per sequence GC content was normally distributed with a mean of 43% and overlapped with the computed theoretical distribution that one would expect
- The SNPs detected for within host diversity assessments did not exceed 4%
- 10 samples had sufficiently high coverage to reconstruct quasispecies

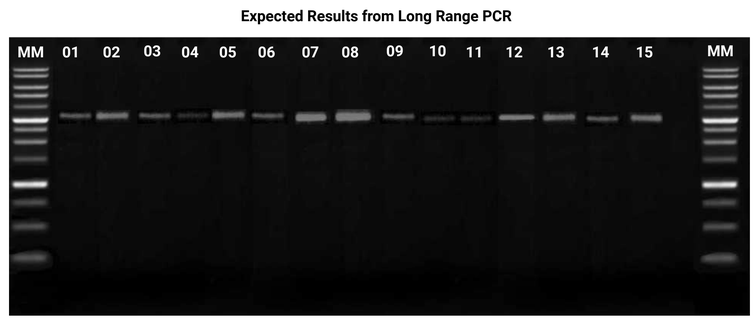


Figure 1: Example of the expected results for a visualised TBE-Agarose gel indicating the molecular marker on either side with the amplicons for samples fully resolved at approximately 3.2 kilo basepair.

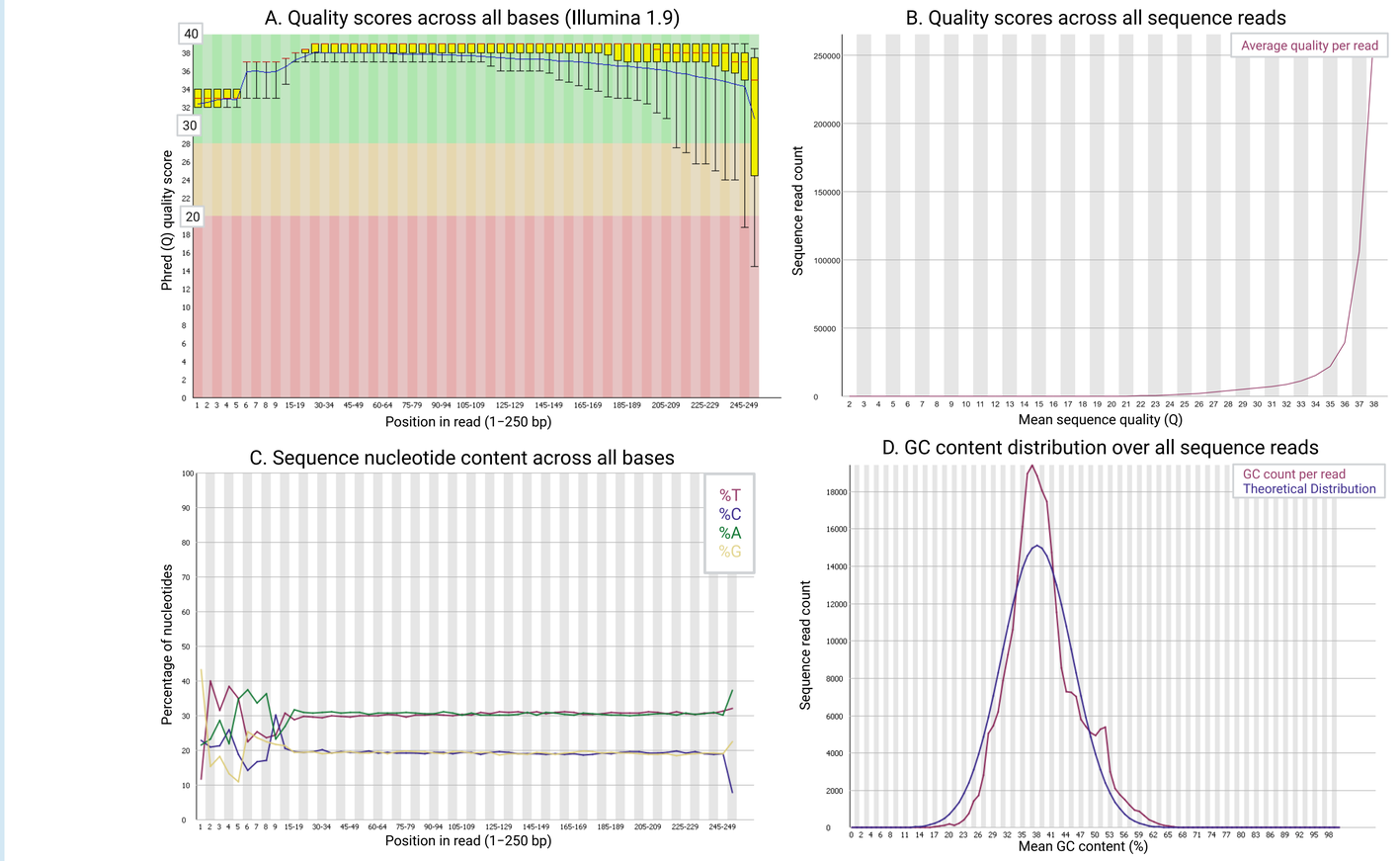


Figure 2: Graphical display of key metrics assessed during the quality control (QC) of read files. A) Plot of average quality scores per base from position 1-250 bp for individual reads. B) Plot of quality scores for reads based on average quality score per read. C) Plot of positional nucleotide composition (%) for reads from position 1-250 bp. D) Comparison of actual versus theoretical distribution of GC content (%).

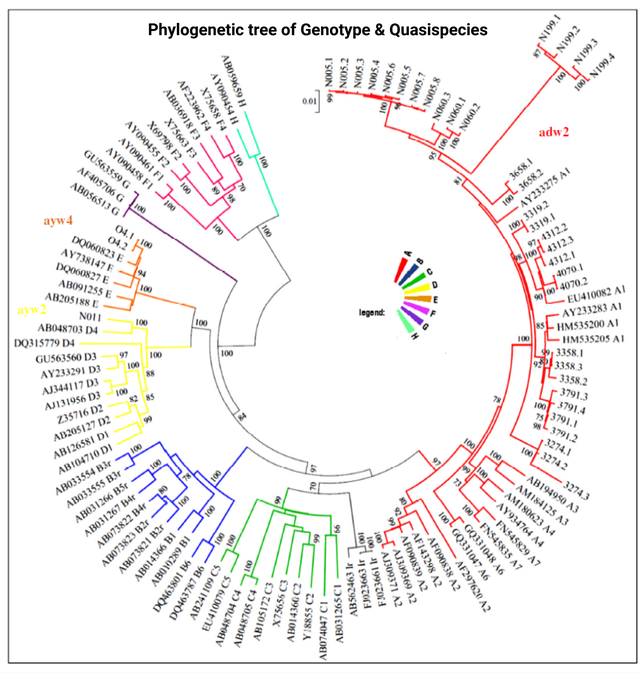


Figure 3: Circular display of phylogenetic tree constructed using the Neighbour-Joining (NJ) method for genotypes A to H, including the reconstructed quasispecies for study samples.

- A total of 37 full genome variants were reconstructed from 10 samples
- Reconstructed quasispecies for individuals ranged from a single major variant to as many as 8 variants
- Larger quasispecies typically consisted of one major variant, a few intermediate variants, and a remaining pool of low frequency variants
- Pairwise differences between variants did not exceed 4%, meaning all variants represented a single genotype and subgenotype
- 8 samples were of genotype A, 1 of genotype D and 1 of genotype E
- Phylogenetic analysis showed that all specimens partitioned within clades of previously described isolates from Africa
- Due to high similarity within individual quasispecies, variants from a single sample formed distinct clusters within their clades

Research funded partially by grants from the National Research Foundation (Grant: 82831) and the Poliomyelitis Research Foundation (Grant: 12/41 MSc)