Comparison of Next-Generation Sequencing and Clone-Based Sequencing in Analysis of Hepatitis B Virus Reverse Transcriptase Quasispecies Heterogeneity

Ling Gong, Yue Han, Li Chen, Feng Liu, Pei Hao, Jia Sheng, Xin-Hua Li, De-Min Yu, Qi-Ming Gong, Fei Tian, Xiao-kui Guo, Xin-Xin Zhang

Department of Infectious Diseases, Institute of Infectious and Respiratory Diseases, Ruijin Hospital, Shanghai Jiaotong University School of Medicine, Shanghai, China; Shanghai Center for Bioinformation Technology, Shanghai, China; Department of Medical Microbiology and Parasitology, Institutes of Medical Sciences, Shanghai Jiaotong University School of Medicine, Shanghai, China

We previously reported that, based on clone-based sequencing (CBS), hepatitis B virus (HBV) heterogeneity within the reverse transcriptase (RT) region was a predictor of antiviral efficacy. Here, by comparing ultradepth pyrosequencing (UDPS), i.e., next-generation sequencing (NGS), with CBS in characterizing the genetic heterogeneity of HBV quasispecies within the RT region, we evaluated the performance of UDPS in the analysis of HBV viral populations. HBV genomic DNA was extracted from serum samples from 31 antiviral treatment-naive patients with chronic hepatitis B. The RT region quasispecies were analyzed in parallel using CBS and UDPS. Characterization of quasispecies heterogeneity was conducted using bioinformatics analysis. Quasispecies complexity values were calculated with the formula $S_n = -\sum (p_{in})/\ln n$. The number of qualified strains obtained by UDPS was much larger than that obtained by CBS ($P < 0.001$). Pearson analysis showed that there was a positive correlation of quasispecies complexity values at the nucleotide level for the two methods ($P < 0.05$), while the complexity value derived from UDPS data was higher than that derived from CBS data ($P < 0.001$). Study of the prevalences of variations within the RT region showed that CBS detected an average of $9.7 \pm 1.1$ amino acid substitutions/sample and UDPS detected an average of $16.2 \pm 1.4$ amino acid substitutions/sample. The phylogenetic analysis based on UDPS data showed more genetic entities than did that based on CBS data. Viral heterogeneity determination by the UDPS technique is more sensitive and efficient in terms of low-abundance variation detection and quasispecies simulation than that by the CBS method, although imperfect, and thus sheds light on the future clinical application of NGS in HBV quasispecies studies.

Hepatitis B virus (HBV) is a noncytopathic DNA virus that infects approximately 350 million people worldwide and is a leading cause of liver cirrhosis and hepatocellular carcinoma. The evolutionary dynamics of HBV are characterized by high mutation rates ($>10^{-5}$ nucleotide substitutions/site/year) due to the error-prone reverse transcriptase (RT) (1) and rapid replication rates ($>10^{4}$ virions per day) (2, 3). Thus, a large number of viral variants are constantly generated, creating great genetic diversity on which natural selection operates. Heterogeneous populations of viruses are often referred to as viral quasispecies (4, 5). The massive heterogeneity present in HBV quasispecies has important biological consequences; it has been identified as an important factor in relation to the clinical features of the infection (transmission, persistence, and liver damage), and it may influence the outcomes of treatment and be important in the development of resistance to nucleoside/nucleotide analogue (NA) antiviral therapy (6).

Cloning of PCR products and subsequent Sanger dideoxy sequencing have been widely used in analyses of the heterogeneity of viral quasispecies and determination of antiviral-associated drug resistance mutations, which has been used as the reference technique to study viral populations. However, this procedure is time-consuming, laborious, and costly, only a few researchers have studied viral populations in considerable detail, and no more than 100 clones were obtained within a patient sample (7). Furthermore, preferential selection of deficient viral genomes can bias the results of molecular cloning (8). Thus, we might have missed many aspects of the viral population based on the dominant sequences, and the full intrapopulation heterogeneity is difficult to characterize. Next-generation sequencing (NGS) technologies have changed the situation dramatically. They allow for massive parallel picoliter-scale amplification and detection of individual DNA molecules (9, 10), which has made it possible to generate hundreds of thousands of clonal sequence reads, providing the potential to reduce the time and complexity for DNA sequencing without the need for cloning. To date, ultradepth pyrosequencing (UDPS), as commercialized with the Roche 454 sequencing platform, has been largely applied in viral research because of the longer read lengths. Typical applications of UDPS in virology are the discovery of unknown and unexpected viruses by metagenomic-based strategies (11), investigation of human immunodeficiency virus (HIV) viral genome variability and human cytomegalovirus (HCMV) genotyping (12, 13), monitoring of antiviral drug resistance in patients with HIV infection (14), and modeling of hepatitis C virus (HCV) evolution within the host (15, 16). NGS can detect low-abundance variations, providing substantial information on the structure...
of the viral population. Understanding the relationship between phenotypic features of the viral population and its genetic structure will help us study the pathogenicity of virus infection and may lead to therapeutic discoveries. Currently, a few applications of UDPS in HBV research have characterized genetic HBV sequence variations (17–20), but none has focused on the HBV quasispecies complexity within the entire RT region on the basis of a data set derived with the UDPS technique. Furthermore, few studies have explored the difference in performance in the analysis of HBV RT gene heterogeneity with the clone-based sequencing (CBS) and NGS methods.

Our previous studies demonstrated the dynamic changes of HBV quasispecies within the RT region during early stages of antiviral therapy in patients who underwent NA treatment using clone-based sequencing (21, 22). In the present study, HBV quasispecies within the RT region were analyzed with GS-FLX standard sequencing technology, which allows sequencing of DNA fragments exceeding 400 bp in length, including the fusion primers, from the same samples as in our previous studies. We compared NGS with CBS in characterizing the genetic heterogeneity of HBV quasispecies within the RT region and evaluated the performance of NGS in the analysis of HBV viral populations in the hope of simplifying and optimizing the procedure of viral heterogeneity determination.

MATERIALS AND METHODS

Patients. The study patients were from two of our published studies, including (i) seven chronic hepatitis B (CHB) patients previously included in a quasispecies study conducted by Chen et al. (21) from our research unit, who had taken lamivudine as a control group in a pegylated alpha-2a interferon phase III clinical trial performed in our hospital in 2002 (23, 24), and (ii) 24 CHB patients who received entecavir monotherapy in our medical centers between 2007 and 2009 and were studied previously by Liu et al. (22) from our research unit. For each patient, collected plasma samples had been stored at −70°C or −20°C. Samples were selected if they showed plasma HBV DNA levels of >100,000 IU/ml at baseline. All patients were treatment naive prior to antiviral therapy. All of them were HBsAg positive for at least 6 months and had elevated serum alanine aminotransferase (ALT) levels before treatment. Exclusion criteria included hepatitis C virus (HCV) or human immunodeficiency virus (HIV) coinfection, liver cirrhosis, and alcohol or drug abuse.

Written informed consent was obtained from all patients, and the study was approved by the ethics committee of the Ruijin Hospital, in accordance with the Helsinki Declaration.

Liver biochemistry, HBV serology, and HBV DNA tests. The liver biochemical analysis was performed with a routine automated analysis system (Beckman Coulter, Fullerton, CA) at baseline. HBV serological markers, including HBsAg, antibody to HBsAg, HBeAg, antibody to HBeAg, and antibody to HBcAg were determined at baseline. All patients were treatment naive prior to antiviral therapy. All of them were HBsAg positive for at least 6 months and had elevated serum alanine aminotransferase (ALT) levels before treatment. Exclusion criteria included hepatitis C virus (HCV) or human immunodeficiency virus (HIV) coinfection, liver cirrhosis, and alcohol or drug abuse.

Molecular cloning and sequencing with the Sanger method. PCR products of 1,096 bp were purified using a QIAquick gel extraction kit (Qiagen), cloned into the pGEM-T vector after the addition of adenylate tails (Promega, Madison, WI), and transformed into TOP 10 Escherichia coli competent cells (Invitrogen) growing on ampicillin plates. Positive clones were identified by PCR assay. Nucleotide sequencing reactions were run on an ABI 3730 automated sequencer (Applied Biosystems, Foster City, CA).

Pyrosequencing. In the UDPS technique, PCR products were purified using a QIAquick gel extraction kit (Qiagen). In order to have a pool of homogeneous samples for downstream pyrosequencing, the concentration of each purified PCR amplicon was first determined using a Nanodrop 2000 spectrophotometer (Thermo Fisher Scientific, Wilmington, DE) according to the manufacturer’s protocol, and the quantity needed to form the sample pool was calculated based on previously determined concentrations. PCR amplicons were mixed at final equimolar concentrations of approximately 0.5 ng/µl and underwent UDPS with the Roche 454 GS FLX platform.

Treatment of sequence data. Raw sequences of clones were aligned to strip gaps using CLUSTAL X, version 1.83 (25). RDP3 software was used to detect potential recombinant sequences (26). To handle the initial UDPS sequence reads, GS Amplicon Variant Analyzer software from Roche was first used to analyze the data and to determine quality scores; reads with low quality scores were discarded. Reads from each individual sample were identified using the sample-specific sequence tags in the primers. To further reduce noise, all reads that contained ambiguous bases or whose lengths lay outside the main distribution, as well as inexact matches to the primer or bar-coding sequence, were removed (27). After stripping of the barcode tags, UDPS reads were aligned to genomic reference sequences from the NCBI (genotype B, GenBank accession no. D00329; genotype C, GenBank accession no. X01587) using the NCBI BLASTN program. Reads whose alignment results covered the wanted region were then collected for downstream analysis.

FIG 1 RT region and overlapped S region of the HBV genome and the positions of amplicons analyzed by NGS. MHR, major hydrophilic region; YMDD, Tyr-Met-Asp-Asp motif.
Detection of heterogeneity of different standard mixtures. In order to compare the heterogeneity determination capacities of CBS and UDPS, plasmid HBV DNA clones harboring the M204V/R mutation were used to generate four mixtures with different proportions of mutant (1%, 10%, 40%, or 70%), the four mixtures were used for CBS and UDPS, and in CBS experiments we chose 20 clones, 40 clones, and 60 clones at each time from each mixture. Then we determined the proportions of mutants in different mixtures.

Sequence analysis. The nature of the viral population was characterized by viral genotypes, genetic complexity, prevalence of variations over sites of the RT region, and phylogenetic analysis using data from the two methods. The HBV genotypes of sequences were determined using the HBV STAR program (28). The quasispecies complexity, known as the Shannon entropy (Sn), was defined as the proportions of different genome sequences in a mutant distribution. The possible values of Sn range from 0 (when all sequences are identical) to 1 (when each sequence is unique). Sn can be calculated with the formula $Sn = -\sum (p_i \ln p_i)/\ln N$, where N is the total number of clones and $p_i$ is the frequency of each clone in the viral quasispecies population (29). The quasispecies complexity was calculated at both the nucleotide level and the amino acid level. The sample sequences were compared with the reference sequences (genotype B, GenBank accession no. D00329; genotype C, GenBank accession no. X01587) to identify variations in the RT region of HBV genomes. Variations from the reference sequences were considered true mutations if they were previously confirmed to confer a phenotypic change in drug susceptibility assays (30); otherwise, they were regarded as possible novel mutations or polymorphisms (18). Established NA resistance-related mutations involve the following RT regions: rtL80I/I, rtL180M, rtA181V, rtT184S/A/E/G, rtS202G/I, rtM204V/I/S, rtN236T, and rtM250V (31). The sensitivity of UDPS in the current platform for detection of low-level viral variants at levels of 0.1% to 1% has been confirmed by the use of standard cloning methods (9, 18, 27, 32). According to previous reports using a restriction target sequence as an internal control, the abundance of variation follows a Poisson distribution, and variants with prevalences of $\geq 1.0\%$ were classified as high-confidence variants. Using a conservative approach to focus on the analysis on the most significant changes, we used this threshold to discriminate authentic variants from possible artifactual errors (17). Phylogenetic trees were constructed using the $p$-distance neighbor-joining method using MEGA 5.1 software (33), and tree topologies were evaluated with a bootstrap value of 2,000 repetitions.

Statistical analysis. The results of continuous variables were expressed as mean ± standard error (SE), and quasispecies complexity values were compared between the two methods using unpaired Student’s t tests. Correlation analysis of the two methods was carried out using the Pearson correlation test. Statistical analyses were performed with statistical software (SPSS version 20.0; SPSS, Chicago, IL). $P < 0.05$ was considered statistically significant.

RESULTS

Demographic, clinical, and laboratory data. Thirty-one patients with CHB were enrolled in the present study. The patients had a mean age of 37.1 ± 2.1 years, there were 25 male and 6 female patients, 14 patients were HBeAg positive, and 17 patients were HBeAg negative. The patients had an average viral load of 7.8 ± 0.2 log$_{10}$IU/ml and an average ALT level of 149.5 ± 16.0 IU/liter.

Comparison of quasispecies strains and genotypes obtained with the two methods. An average of 25.2 ± 0.2 validated quasispecies strains per sample were generated by the clone sequencing method, and a total of 780 analyzable clones were obtained. The UDPS approach, before being mapped to HBV reference strains, yielded an average of more than 1,000 (up to 10,000) sequences reads. After being denoised and mapped, the number of “reads,” i.e., the number of qualified quasispecies, was on average 1,120 ± 123.5 strains per patient for the first RT region, 928 ± 60.7 strains per patient for the second RT region, and 545.5 ± 32.9 strains per patient for the third RT region. The number of qualified strains obtained by UDPS was much larger than that obtained by CBS in all three overlapping RT regions ($P < 0.001$ (Fig. 2). HBV genotypes determined using the data from the two methods were completely matched. Eleven patients were infected with genotype B HBV strains and 20 with genotype C strains.

Heterogeneity detection capacity of CBS and UDPS. The results are shown in Table 1. The UDPS method was able to detect mutant strains down to 1%, while CBS could not. Overall results showed that the mutant percentages determined by the UDPS method were closer to the expected percentages than those determined by CBS.

Correlation and comparison of quasispecies complexity values derived with the two methods. Correlations of quasispecies complexity values derived with the two methods were determined at the nucleotide level and the amino acid level for all three overlapping RT regions. At the nucleotide level, the complexity value derived with the UDPS method was correlated positively with that derived with the CBS method for all three overlapping RT regions (RT region 1, $r = 0.624$, $P < 0.001$; RT region 2, $r = 0.562$, $P < 0.001$; RT region 3, $r = 0.350$, $P < 0.05$). At the amino acid level, the complexity value derived with the UDPS method was correlated positively with that derived with the CBS method for the first and second RT regions (RT region 1, $r = 0.512$, $P < 0.01$; RT region 2, $r = 0.419$, $P < 0.05$), but there was no significant correlation between the two methods for the third region ($P > 0.05$). A

![FIG 2 Numbers of qualified strains (QS) obtained with the two methods in three overlapping RT regions. ***, $P < 0.001$.](http://jcm.asm.org/ Downloaded from)

<table>
<thead>
<tr>
<th>Expected mutant %</th>
<th>Mutant % observed:</th>
</tr>
</thead>
<tbody>
<tr>
<td>CBS</td>
<td>20 clones</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>10</td>
<td>5</td>
</tr>
<tr>
<td>40</td>
<td>25</td>
</tr>
<tr>
<td>70</td>
<td>60</td>
</tr>
</tbody>
</table>
A dot scatter plot of paired HBV RT quasispecies complexity values determined by CBS and UDPS is shown in Fig. 3.

Complexity values calculated using the data derived from the two methods were also compared. As expected, viral quasispecies complexity values generated with the UDPS technique were higher than those determined by the CBS method for all overlapping RT regions at both the nucleotide level and the amino acid level \( (P < 0.001) \) (Fig. 4 and Table 2).

Comparison of the prevalences of variations over RT region sites. The prevalences of variations over sites of the RT region, including NA resistance-related mutations, among quasispecies derived with these two methods were compared. The frequencies

![FIG 3](image_url)  
**FIG 3** Dot scatter plot of paired HBV RT quasispecies complexity values determined by CBS and UDPS \( (n = 31) \). *, \( P < 0.05 \); **, \( P < 0.01 \); ***, \( P < 0.001 \).

![FIG 4](image_url)  
**FIG 4** HBV RT quasispecies complexity values based on CBS \( (n = 31) \) and UDPS \( (n = 31) \) data. ***, \( P < 0.001 \). The line above the box indicates the maximum value. The upper line of the box indicates the 75th percentile. The line in the box indicates the median. The lower line of the box indicates the 25th percentile. The line below the box indicates the minimum value.
of amino acid substitutions, including NA resistance-related mutations, detected with the two methods are shown in Table S1 in the supplemental material. The overall distributions of nucleotide and amino acid substitutions within the RT region detected with the two methods for one patient (patient 29) are shown in Fig. 5. As expected, UDPS determined more amino acid substitutions and more NA resistance-related mutations in the RT region, which made UDPS results more informative than results obtained with the CBS method.

CBS detected an average of 9.7/1.1 amino acid substitutions/sample, and UDPS detected a mean of 16.2/1.4 amino acid substitutions/sample. Among these substitutions, CBS identified a mean of 2.3/0.2 more substitutions/sample that were not detected by UDPS, and UDPS identified a mean of 8.8/0.7 more substitutions/sample that were missed by CBS.

Among the NA resistance-related mutations, both the UDPS and CBS methods identified A181T and M204I in 3 samples, CBS identified L180M in 1 sample alone, and UDPS identified A181T, M204I, and M250V in 8 samples alone. L180M was present at a frequency of 10% and was detected only by the CBS method for patient 16. A181T was present at frequencies ranging from 1.2% to 5.7% and was detected for 2 patients (patients 26 and 29) by both methods and for 4 patients (patients 12, 25, 28, and 30) only by the UDPS method. M204I was present at frequencies ranging from 1.7% to 4.7% and was detected for patient 16 by both methods and for 3 patients (patients 24, 25, and 31) only by the UDPS method. M250V was present at a frequency of 10% and was detected only by the UDPS method for patient 7.

**Comparison of quasispecies phylogenetic simulations.** Phylogenetic trees were generated from RT region 2 in one patient (patient 29) (Fig. 6). Some parts of the two topologies resembled each other, while the structure topology of the tree constructed from the UDPS data was more complex than that constructed from the CBS data, and the branch lengths of the tree constructed from the UDPS data were more diverse than in that constructed from the CBS data. The phylogenetic analysis based on the UDPS data showed more genetic entities than did that based on the CBS data. In this study, CBS methods generated an average of 25 clones per sample, while UDPS yielded up to nearly 1,000 validated strains that could be used for phylogenetic study.

**DISCUSSION**

The complex “cloud-like” nature of viral quasispecies allows them to evolve rapidly and to adapt continuously to new environments by selecting most-fit distributions (34). The population size at-

**TABLE 2 Quasispecies complexity values based on CBS and UDPS methods**

<table>
<thead>
<tr>
<th>Level and region</th>
<th>Complexity value for:</th>
<th>CBS</th>
<th>UDPS</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>RT 1</td>
<td>RT 2</td>
<td>RT 3</td>
</tr>
<tr>
<td>Nucleotide</td>
<td></td>
<td>0.3678 ± 0.0345</td>
<td>0.6742 ± 0.0211</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.4503 ± 0.0315</td>
<td>0.5873 ± 0.0312</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.4581 ± 0.0341</td>
<td>0.7437 ± 0.0271</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Amino acid</td>
<td></td>
<td>0.1785 ± 0.0234</td>
<td>0.4767 ± 0.0172</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.2719 ± 0.0308</td>
<td>0.4631 ± 0.0295</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.2370 ± 0.0291</td>
<td>0.6534 ± 0.0304</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>

**FIG 5** Overall distributions of nucleotide and amino acid substitutions and their relative frequencies within the RT region of HBV detected with the two methods for patient 29.
tained by a virus in an infected host can be exceedingly large, which may affect the evolutionary outcome (35). Thus, knowledge of the large amount of genetic information in viruses is required to explain the properties of viruses. Until recently, cloning of PCR products has been widely used as the reference technique to study the genetic heterogeneity of viral populations; however, its sensitivity is limited by the number of clones that can be feasibly sequenced. An average of 25 clones per patient was obtained in this study, which represented a very tiny portion of the actual viral population in the patient. In this respect, UDPS technology is superior to traditional CBS methods in its ability to generate a large amount of sequence data in one instrument run, achieving hundreds-fold greater coverage with ease. Indeed, after filtering and mapping strategies, UDPS generated up to 1,000 reads per patient in this study, including sequences from dominant, intermediate, and minor viral populations, providing a snapshot of the entire viral population and resulting in a much larger quasispecies population size than with cloned PCR products.

Heterogeneity determinations indicated that the CBS method had more bias than the UDPS method. Thus, sufficient viral quasispecies strains provide the basis for assessing viral diversity, avoiding either reduction or inflation of the true viral heterogeneity. This gave the UDPS-based complexity calculations greater capacity in terms of simulation of the real viral quasispecies heterogeneity. In this research, UDPS generated higher viral quasispecies complexity values for all of the overlapping RT regions than did CBS, at both the nucleotide level and the amino acid level, and quasispecies complexity values derived with the two methods were correlated at the nucleotide level. We can anticipate that viral complexity calculations with mathematical models based on UDPS data would be more reliable than those with models based on CBS data.

In population genetics studies, the read length of a virus strain should be considered. By simulation of a sample profile with concentrations of 330 bacteria, a previous study showed that the ability to identify genera improves substantially when read lengths are increased beyond 100 bp (36). In our study, with sequencing read lengths of 400 bp, multiple variable regions could be covered by a single read, offering superior phylogenetic identification and taxonomic assignment.

Drug resistance is the principal cause of antiviral treatment failure, which may result in clinical disease progression. CBS is currently a standard approach for detecting minor drug-resistant variants, yet it is not sensitive enough to detect variants present in very small proportions due to their poor relative fitness, compared to wild-type virus. Allele-specific hybridization, another currently standard approach (e.g., INNO-LiPA; Innogenetics), can detect only single mutations at a particular position, and its sensitivity is limited by nonhybridization due to polymorphisms within the region of the probes (37). For the UDPS technique, the sensitivity depends on the final number of genome copies that can be successfully extracted and amplified from a plasma sample; the specificity is limited by the errors occurring during the PCR and UDPS procedures (38). In our study, the HBV DNA load (7 log10 IU/ml, on average) was high in the serum samples, and we optimized DNA extraction protocols to obtain adequate templates. In addition, we used the high-fidelity enzyme Pfx to improve accuracy during PCR amplification. It has been reported previously that insertions and deletions are the most common types of errors in pyrosequencing and occur mainly in homopolymeric regions, where linear relationships between signal intensity and the number of incorporated nucleotides are no longer preserved (9). Several independent studies suggested miscellaneous cutoff values and in-house probabilistic algorithms to reduce noise; these cutoff...
values are generally low, ranging from 0.03% to 0.21% (15, 20, 39). In our study, we applied quality control criteria to reduce noise, based on quality scores or on properties of the reads, and we used a threshold (≥1.0%) for detection of high-confidence variants for the purpose of focusing on the most significant variations. It is possible to improve the sequencing error rate to some extent, but it is impossible to eliminate errors completely (40).

Using CBS and UDPS techniques, we found that a number of amino acid substitutions, both novel and already described as being associated with drug resistance, were already present with variable frequencies in all of the drug-naïve patients, which was similar to previous reports (17, 18). Many mutations may be selectively neutral and thus submit to genetic drift, and their presence may enable the virus to respond more rapidly to changed environments (41). Among the NA resistance-related mutations, A181T and M204R were identified more frequently. It has been reported that G-to-A hypermutation at these codons is responsible for a drug-resistant mutation. The mechanism underlying this phenomenon is not clear, and this could be analyzed further in future studies involving host factors such as APOBEC 3 (42). In our study, all of the mutations detected by CBS were not determined by the UDPS method. The reason underlying this might include the following. (i) Biased amplification might skew frequency estimates of generic variants due to primer mismatches with the two methods during PCR (43). (ii) Insertions/deletions and mismatch errors were common during pyrosequencing, especially in homopolymeric regions. (iii) “Low-quality” reads were discarded because of the rigid filtering strategy used during pyrosequencing. The influence on the clinical significance of low-prevalence NA resistance-related mutations in the NA-naïve CHB patients cannot be ascertained due to the cross-sectional nature of our study. It remains unclear whether these minor drug-resistant mutations have clinical significance. Simen et al. reported that UDPS identified a large proportion of chronically HIV-infected, antiretroviral therapy-naïve patients who harbored drug-resistant variants, and the detection of these previously occult resistant strains predicted virological failure for patients receiving antiretroviral regimens (14). Large-scale studies are required for characterization of the baseline and dynamics of HBV resistance with UDPS data, to guide HBV clinical treatment strategies.

Although UDPS is a promising method for characterizing the genetic heterogeneity of HBV quasispecies, a few disadvantages might limit its extensive application. Insertions/deletions and mismatch errors can occur during PCR amplification and pyrosequencing, and amplification bias might affect the relative frequencies of viral strains. Subsequent data analysis steps require adequate computing resources and qualified data-processing software packages. However, current software programs are relatively new, and more-extensive validation and comparisons are necessary to better understand their performance in the analysis of viral populations.

In conclusion, we give an overview of the application of both CBS and UDPS methods in the analysis of viral populations; our results show that UDPS coupled with analytic methods has superiority in characterizing the genetic heterogeneity of HBV RT quasispecies with enhanced sensitivity and efficiency. The results shed light on the future clinical application of quasispecies complexity determinations and detection of low-abundance variations to direct antiviral treatment for CHB patients. The UDPS approach offers a way to obtain a snapshot of the entire viral population. This snapshot is imperfect, but it can be improved through sample template preparation, optimization of PCR amplification conditions, and data processing. With decreases in costs, improvements in sequencing quality, and increases in read lengths, we are optimistic that NGS can be used to gain new insights regarding HBV evolution, pathogenesis, and drug resistance for translation research.

ACKNOWLEDGMENTS

This study was supported by the National Nature Science Foundation of China (grant 81171616), National Key Research Projects of China (grants 2012ZX10002007, 2012CB519002, and 2013ZX10002001), and the Shanghai International Collaboration Project (grant 11410701700).

REFERENCES


