Published in final edited form as:

Infect Genet Evol. 2013 December ; 20: 225–229. doi:10.1016/j.meegid.2013.08.014.

A panel of 16 full-length HCV genomes was characterized in China belonging to genotypes 1 to 6 including subtype 2f and two novel genotype 6 variants

Ru Xu¹, Wangxia Tong², Lin Gu², Chunhua Li, Yongshui Fu¹, and Ling Lu^{2,3,*}

¹Guangzhou Blood Center, Guangzhou, Guangdong, China

²Hepatology Laboratory, Third Affiliated Hospital of Sun Yat-sen University, Guangzhou, Guangdong, China

³The Viral Oncology Center, Department of Pathology and Laboratory Medicine, University of Kansas Medical Center, Kansas City, Kansas, USA

Abstract

In this study, we characterized the full-length genomes of 16 HCV isolates obtained from patients in a single hospital in China using overlapping PCR followed by DNA sequencing. The obtained genomes are 9,414–9,628 nucleotides in length, and each genome contains a single ORF of 9,021–9,102 nucleotides. Nine genomes represent the common subtypes 1a, 1b, 2a, 2b, 3a, 3b, and 6a, while seven represent the infrequent lineages 1c, 2f, 4d, and 5a, and two novel genotype 6 variants. GZ51969 and GZ52540 are subtype 1b isolates belonging to two unique clusters designated A and B, which account for 29.5% and 59.5% of the 1b infections in China, respectively. ZS542 and GZ98799 represent the first two complete genomes of the provisionally assigned subtype 2f. ZS96 and ZS202 are novel genotype 6 variants that may qualify for two new subtypes. ZS17, ZS537, and ZS631 represent three alien subtypes, namely, 1c, 4d, and 5a, which were detected in China for the first time in this study and may have been recently introduced as a result of globalization. Taken together, these results confirmed a large variety of HCV taxonomic lineages in China through the sequencing of their full-length genomes. These lineages represent six genotypes, 11 subtypes, and two novel variants. They were characterized for achieving a better understanding of the HCV genetic variation patterns and for possible future research applications.

Keywords

HCV; full-length genome; genotype; subtype; China

© 2013 Elsevier B.V. All rights reserved.

*Correspondence: Ling Lu, The Viral Oncology Center, Department of Pathology and Laboratory Medicine, University of Kansas Medical Center, MSN 3041, 3901 Rainbow Boulevard, Kansas City, Kansas City, KS 66160. llu@kumc.edu. Tel: 913-945-6451 Fax: 913-945-6836.

Xu R and Tong WX contributed equally to this study.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

1. Introduction

The analysis of genetic sequences has resulted in the classification of the hepatitis C virus (HCV) into six confirmed and one provisional genotype, while within each genotype except for genotypes 5 and 7 a number of subtypes have been designated (Kuiken et al., 2008). Based on the nucleotide identity, a genome-wide difference of 31%-33% is sufficient to distinguish genotypes and a genome-wide difference of >15% is sufficient to distinguish subtypes (Simmonds et al., 2005). HCV genotypes and subtypes exhibit different geographical distribution patterns (Simmonds, 2001). In general, subtypes 1a, 1b, 2a, 2b, and 3a are found worldwide, whereas the others are restricted to particular regions (Simmonds et al., 2005; Pybus et al., 2007; Pybus et al., 2009). However, such patterns are constantly changing as a result of modern transmission and global travel.

The provisional assignment of HCV genotypes and subtypes is based on the analysis of partial sequences in the E1 and NS5B regions (Simmonds et al., 2005). Using this strategy, we have determined the HCV genotype distribution patterns in different regions in China and between different population subsets. Overall, subtype 1b is predominant followed by 2a. However, in the south 6a has replaced 2a as the second most predominant subtype and represents the emerging HCV strains among volunteer blood donors and IDUs (Lu et al., 2005; Xia et al., 2008a; Fu et al., 2011; Fu et al., 2012; Gu et al., 2013). Other HCV genotypes and subtypes have also been detected. These include 1a, 2b, 3a, 3b, 6e, and a few isolates of 1c, 2f, 4d, 5a, and some novel genotype 6 variants. Of note, we have recently detected all of these HCV taxonomic lineages in patients sampled in a single hospital in Guangdong Province, and this finding demonstrates the high genetic complexity of HCV found in China (Gu et al., 2013). In the present study, we characterized the full-length genomes of 16 isolates to further analyze this genetic complexity.

2. Materials and Methods

2.1. Subjects and specimens

We recently characterized the partial sequences of the E1 and NS5B regions of HCV in 393 patients in the 3rd Affiliated Hospital of Sun Yat-sen University in China (Gu et al., 2013). Thirteen isolates from that study (with the initial of ZS before a number) were selected for full-length genome sequencing in the present study. In an earlier report, we analyzed the HCV genotype distribution patterns in China (Lu et al., 2005), and three isolates (GZ51969, GZ52540, and GZ98799) from that report were also selected for entire genome sequencing in the present study because they represent the 1b strains in the A and B clusters and a strain of the provisionally proposed subtype 2f (Table 1). The three isolates were also sampled in the 3rd Affiliated Hospital of Sun Yat-sen University in China. In total, samples from 16 patients were used in the present study. Of these, ZS537 was obtained from a European visitor, from which the subtype 4d isolate was characterized. ZS96 was from an overseas Chinese individual living in Myanmar, from which a novel genotype 6 variant was determined. With the exception of these two patients, all of the other patients were native Chinese.

2.2. PCR amplification and sequence analysis

Each full-length HCV genome was sequenced in 100 µl of serum using previously described approaches (Lu et al., 2005). To avoid possible carryover contamination, standard procedures were used (Kwok and Higuchi, 1989). The obtained HCV sequences were then analyzed using the BioEdit software (Tippmann, 2004). Prior to phylogenetic tree reconstruction, the best-fitting substitution model was selected using jModeltest based on the Akaike Information Criterion (Posada and Crandall, 1998). Consistent with our recent results, GTR+I+ Γ_6 was found to be the best for all of the datasets (Fu et al., 2012). Under this model, maximum likelihood (ML) phylogenetic trees were heuristically searched using the subtree pruning and regrafting (SPR) algorithm and the nearest neighbor interchange (NNI) perturbation algorithm implemented in PHYL (Guindon and Gascuel, 2003). After the tree files were generated, the tree topology was displayed, and the nucleotide similarities were calculated using the MEGA5 software (Kumar et al., 2004). To detect possible virus recombination events, we used RDP3 as previously described (Martin et al., 2010; Li et al., 2012).

2.3. Nucleotide Sequence Accession Number

The nucleotide sequences reported in this study were deposited into GenBank with the following accession numbers: KC844039 through KC844052.

3. Results

3.1. Determination of full-length HCV genomes

Full-length HCV genomes were determined each from 13-28 overlapping fragments for the 16 isolates outlined in table 1. The obtained genomes were 9414 to 9628 nucleotides (nt) in length. The 5'UTRs were 298 to 341 nt in length, whereas the 3'UTRs ranged from 26 to 258 nt in length. Seven isolates (ZS30, ZS96, ZS202, ZS221, ZS574, ZS633, and ZS674) had their 3'UTRs amplified through to the poly(U) tracts, and two isolates (GZ51969 and GZ52540) had their 3'UTRs amplified through to the 3'X tails. However, the poly(U) tracts of six of the isolates (ZS17, ZS260, ZS542, ZS631, ZS633, and ZS650) were not amplifiable, and the 3'UTR of one of the isolates (ZS620) was not obtained. The 16 isolates each contained a single ORF that was 9021 to 9102 nt in length, and the longest ORFs were found in the ZS26, ZS542, ZS623, and GZ98799 isolates, and the shortest ORF was found in ZS537. The predicted 10 HCV proteins had the following sizes (Table 1): core, 573 nt/191 aa; E1, 576 nt/192 aa; E2, 1086-1113 nt/362-371 aa; P7, 189 nt/63 aa; NS2, 651 nt/217 aa; NS3, 1893 nt/631 aa; NS4A, 162 nt/54 aa; NS4B, 783 nt/261 aa; NS5A, 1332-1398 nt/444-466 aa; and NS5B, 1776 nt/591 aa.

3.2. Pairwise comparison of nucleotide sequences

Of the 16 isolates, 12 belong to subtypes that had their prototypic full-length genomes characterized (Figures 1-3) (Simmonds et al., 2005). For this reason, these 12 isolates are not described in this section, and the remaining four isolates (ZS96, ZS202, ZS542, and GZ98799) were examined. In our recent studies based on partial sequences, ZS542 and GZ98799 were classified into subtype 2f, whereas ZS96 and ZS202 represented two novel genotype 6 variants (Lu et al., 2005; Gu et al., 2013). Accordingly, we compared the

nucleotide similarities of ZS542 and GZ98799 with 16 sequences of genotype 2 (representing 2a-2e, 2i-2k, 2m, 2r, and six equivalents) and the similarities of ZS96 and ZS202 with 28 sequences of genotype 6 (representing subtypes 6a-6w and five equivalents).

To date, two isolates have been assigned into subtype 6w, but they only had incomplete ORF sequences characterized: D140 (EU643834) and D370 (EU643836). In one of our previous studies, we determined the full-length genome for a novel genotype 6 variant, namely GZ52557 (DQ278892), which can also be classified as subtype 6w. However, in the Los Alamos HCV database, this isolate is listed as an unclassified genotype 6 variant because it was deposited before the designation of subtype 6w was proposed (Lu et al., 2006). To include a full-length ORF sequence for analysis in this study, we used GZ52557 to represent subtype 6w.

Compared with each other, ZS542 and GZ98799 exhibited nucleotide similarities of 90.8% across the whole genome and 90.4% across the entire ORF. These similarities indicate that they both are distinct isolates of a single subtype. Compared with the 16 genotype 2 sequences, ZS542 and GZ98799 presented nucleotide similarities of 74.4%-84.4% across the whole genome and 76.7%-84.4% across the entire ORF (Supplementary Table 1). These similarities indicate that they both represent a subtype that was sequenced in its entirety for the first time in this study. Compared with each other and with the 28 genotype 6 sequences, ZS96 and ZS202 exhibited nucleotide similarities of 72.6%-76.8% across the whole genome and 71.9%-76.1% across the entire ORF (Supplementary Tables 2-3). Because these similarities are all above the subtype level, ZS96 and ZS202 represent distinct variants equivalent to new subtypes.

3.3. Phylogenetic analysis

Three ML trees were reconstructed based on the full-length HCV genome sequences. These trees showed that ZS30 was grouped into subtype 1a, both GZ51969 and GZ52540 were grouped into 1b, and ZS17 was grouped into 1c. In addition, ZS623 was grouped into subtype 2a, ZS260 was grouped into 2b, ZS633 was grouped into 3a, ZS650 was grouped into 3b, ZS537 was grouped into 4d, ZS631 was grouped into 5a, and both ZS221 and ZS674 were grouped into 6a (Figure 1-3). Because these 12 isolates each showed genetic distances of <15% to their nearest relatives, their subtype classification was confirmed (Simmonds et al., 2005). Figure 2 presents the ML tree based on 21 full-length genome sequences of genotype 2, representing 12 subtypes and six equivalents. This ML tree shows that ZS542 and GZ98799 formed a subset that is distinct from all of the other genotype 2 lineages but adjacent to the QC259/2d branch. Their distances to QC259/2d are >15%, which is sufficient to confirm the designation of ZS542 and GZ98799 as subtype 2f. Figure 3 presents the ML tree based on 37 full-length genome sequences of genotype 6, representing 23 assigned subtypes and five equivalents. Although not all of the subtypes in this tree were well separated, ZS96 and ZS202 were recognized as two distinct subtype equivalents. ZS202 was grouped close to subtypes 6a and 6b, whereas ZS96 was clustered with subtypes 6u and 6v. In one of our previous studies, subtypes 6u and 6v represented a distinct subset that was linked to the tree by twin branches, each with a long genetic distance

(Xia et al., 2008a). It was shown in the present study that this subset now has an intermediate relative, namely ZS96 (Figure 3).

3.4. Similarity plotting

It is possible that the phylogenetic tree topology may have been affected by recent viral recombination events. To exclude this possibility, pairwise similarity scores were calculated by individually comparing ZS542 and GZ98799 with 16 genotype 2 sequences (Supplementary Table 1) and ZS96 and ZS202 with 28 genotype 6 sequences (Supplementary Tables 2 and 3). In each case, the comparisons with the reference sequences resulted in roughly similar curve patterns. Except for the known highly variable regions in E2 and NS5A, no additional acute changes in the similarity curves were observed. Because other recombination detections also yielded negative results, the possibility of recent major recombination events was excluded (data not shown).

4. Discussion

In this study, 16 full-length HCV genomes were characterized from patients sampled in a single hospital. These genomes represent six genotypes (genotypes 1 to 6), 11 subtypes (1a, 1b, 1c, 2a, 2b, 2f, 3a, 3b, 4d, 5a, and 6a), and two unclassified variants (ZS96 and ZS202). This is the first report to demonstrate the great complexity of HCV taxonomic lineages in China by providing a panel of full-length genome sequences for representing the epidemic and endemic strains. It is not surprising that subtypes 1b, 2a, 3a, 3b, 6a, and 6e were detected because they are prevalent HCV lineages in China (Lu et al., 2005; Xia et al., 2008a; Fu et al., 2011; Fu et al., 2012; Gu et al., 2013). However, it is surprising that five rare subtypes, namely 1c, 2b, 2f, 4d, and 5a, and two novel genotype 6 variants were also detected.

In one of our recent studies, the full-length HCV genome sequences were determined for 11 isolates of genotype 2 sampled from African, European, and Southeast Asian immigrants living in Quebec, Canada. These 11 isolates represent five assigned subtypes (2d, 2e, 2j, 2m, and 2r) and six unclassified equivalents (Li et al., 2012). In combination with those archived in the Los Alamos HCV database, a total of 10 assigned subtypes have since been confirmed for genotype 2 (Kuiken et al., 2008). In the present study, the full-length genomes were determined for one more genotype 2 subtype, namely 2f, through the sequencing of the isolates ZS542 and GZ98799, which were sampled in China. In the consensus paper on HCV nomenclature, 2f was categorized as a “provisionally assigned subtype” due to the absence of a full-length genome sequence (Simmonds et al., 2005). The results from the present study confirmed subtype 2f.

In two other studies, we determined the full-length genomes of six variants of genotype 6: DH012, DH014, DH028, KM046, KM181, and KMN02. These variants exhibit considerable genetic differences from the reference isolates. Based on the criteria of the updated HCV nomenclature, we have assigned these into two new subtypes: 6u and 6v. Phylogenetically, these subtypes were represented by two clusters of tightly grouped isolates, and each cluster was linked to the phylogenetic tree through a long branch (Xia et al., 2008b; Wang et al., 2009). With the addition of the full-length ZS96 genome that was

characterized in the present study, these two clusters now have a third sibling. To demonstrate this finding, an extensive phylogeny was plotted, and this phylogeny revealed that all of the HCV genotype 6 lineages can be divided into two clades and an “insufficient” subset (i.e., 6u, 6v, and ZS96). Because the two clades each contain numerous subtypes (i.e., subtypes 6a-6g, 6o-6t, and 6w were included in one clade, and subtypes 6h-6n and nine novel variants were contained in the other clade), a history of active transmission and divergence was suggested. However, the two clades are in sharp contrast to the “insufficient” subset, for which either a unique evolutionary pattern was involved or many members remain missing, but we tend to believe the latter. Geographically, subtypes 6u and 6v have been identified in Myanmar and in a province of China neighboring Myanmar (Xia et al., 2008b; Wang et al., 2009). Consistently, the newly identified ZS96 isolate was isolated from an overseas Chinese individual living in Myanmar (Gu et al., 2013). Collectively, these findings suggest that the “insufficient” subset may only include the genotype 6 isolates that are indigenous to Myanmar. However, due to the absence of an extensive HCV surveillance program in Myanmar, we are currently unable to confirm this hypothesis.

Historically, subtypes 6a and 6b were discovered prior to any other genotype 6 lineages. In the present study, this subset became slightly complicated after the novel variant ZS202 was included. Similarly to the above-described subset containing 6u and 6v, this finding may also suggest the possibility of missing members. We have identified new genotype 6 variants that cluster within the 6a, 6b, and ZS202 subset. Although their full-length genome sequencing is currently under completion, these variants appear to obscure the clear separation between 6a and 6b (data not shown). Unlike 6u and 6v and the ZS96 isolate, these newly identified variants were exclusively detected in Laos (Wang et al., 2013). However, due to the lack of further information for the patient from whom ZS202 was isolated, we are uncertain whether ZS202 is also epidemically linked to Laos (Gu et al., 2013).

In terms of epidemiology, some subtype 1b isolates are unique to China and are genetically distinct from those found in other countries. Of them, two major clusters (A and B) have been tentatively classified. These clusters account for 29.5% and 59.5% of the total 1b infections in China, respectively, and show different geographic distribution patterns. In general, cluster A is prevalent nationwide, whereas cluster B is predominant in Guangdong Province (Lu et al., 2005; Gu et al., 2013). In this study, we describe the full-length genomes of two HCV isolates belonging to these two clusters: GZ52540 belonging to cluster A and GZ51969 belonging to cluster B. We hope that these data will contribute to the understanding of the molecular epidemiology of genotype 1b isolates in China and the world.

The search of the Los Alamos HCV database (<http://hcv.lanl.gov>, accessed on 05/13/2012) for isolates with known sampling countries identified a total of 42 full-length or nearly full-length HCV genome sequences that originated in China (Kuiken et al. 2008). These represent four genotypes (1, 2, 3, and 6) and 10 subtypes (1b, 2a, 3a, 6a, 6e, 6k, 6n, 6u, 6v, and 6w; Supplementary Figure 1). In this study, we provided such information for an additional seven subtypes, namely 1a, 1c, 2b, 2f, 3b, 4d, and 5a, and two novel genotype 6

variants. As a result, full-length genome sequences are now available for 58 isolates in China, which represent six genotypes, 18 assigned subtypes, and a number of unclassified variants and thus confirm the marked genetic diversity of HCV in China.

Two uncertainties were discussed. The first one concerns the geographic regions where the patients acquired the HCV infections with subtype 1c, 2b, 2f, 4d, 5a, and two novel variants because these HCV lineages do not appear to be frequently encountered in China. With the exception of the 4d isolate, which was obtained from a European visitor, and a novel variant that was obtained from a Chinese individual living in Myanmar, all of the other isolates were obtained from native Chinese. Of the latter, three possibilities may be responsible: (1) these patients may have journeyed to countries where these HCV strains are circulating; (2) due to rapid globalization, an increasing number of foreigners visited China and might have introduced these alien strains; and (3) certain of these strains may have been endemic to niche areas in China for a long time, and the accelerated economic development has opened these areas to the public. Unfortunately, due to the lack of detailed patient information associated with the samples used in this study, we cannot determine which possibility is true. The second uncertainty concerns the subtype designation of the two novel genotype 6 variants. The analysis of their full-length genomes supports the hypothesis that they each are equivalent to a new subtype; however, a formal designation was not provided due to the absence of each three closely related isolates. After the inclusion of all partial genotype 6 sequences in GenBank for further analyses, no additional highly similar isolates were revealed (data not shown). Due to the lack of extensive HCV epidemiological studies in most Southeast Asian countries, it is difficult to exclude the possibility that these two variants represent HCV strains associated with local epidemics in certain regions in these countries or in China. According to the criteria in the consensus paper on HCV nomenclature, now we can only categorize these isolates as two unassigned genotype 6 variants until additional highly similar isolates are identified (Simmonds et al., 2005).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

5. Financial Disclosure: The present study was supported by a grant to Ling Lu from The National Institute of Allergy and Infectious Diseases (5 R01 AI080734-03A) and a grant to Yongshui Fu from The Science and Information Technology Bureau, Guangzhou Municipality, China (No. 2010UI-E00661). The funding agencies had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

References

- Fu Y, Qin W, Cao H, et al. HCV 6a prevalence in Guangdong province had the origin from Vietnam and recent dissemination to other regions of China: phylogeographic analyses. *PLoS One*. 2012; 7:e28006. [PubMed: 22253686]
- Fu Y, Wang Y, Xia W, et al. New trends of HCV infection in China revealed by genetic analysis of viral sequences determined from first-time volunteer blood donors. *J Viral Hepat*. 2011; 18:42–52. [PubMed: 20196805]

- Gu L, Wang T, Yuan M, et al. An increased diversity of HCV isolates were characterized among 393 patients with liver disease in China representing six genotypes, 12 subtypes, and two novel genotype 6 variants. *J Clin Virol.* 2013; 57:311–317.
- Guindon S, Gascuel O. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol.* 2003; 52:696–704. [PubMed: 14530136]
- Kuiken C, Harber P, Thurmond J, et al. The hepatitis C sequence database in Los Alamos. *Nucleic Acids Res.* 2008; 36:D512–516. [PubMed: 18025038]
- Kumar S, Tamura K, Nei M. MEGA3: Integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment. *Brief Bioinform.* 2004; 5:150–163. [PubMed: 15260895]
- Kwok S, Higuchi R. Avoiding false positives with PCR. *Nature.* 1989; 339:237–238. [PubMed: 2716852]
- Li C, Cao H, Lu L, et al. Full-length sequences of 11 hepatitis C virus genotype 2 isolates representing five subtypes and six unclassified lineages with unique geographical distributions and genetic variation patterns. *J Gen Virol.* 2012; 93:1173–1184. [PubMed: 22357752]
- Lu L, Nakano TY, He Y, et al. Hepatitis C virus genotype distribution in China: predominance of closely related subtype 1b isolates and existence of new genotype 6 variants. *J Med Virol.* 2005; 75:538–549. [PubMed: 15714489]
- Lu L, Nakano T, Li C, et al. Hepatitis C virus complete genome sequences identified from China representing subtypes 6k and 6n and a novel yet unassigned subtype within genotype 6. *J Gen Virol.* 2006; 87:629–634. [PubMed: 16476984]
- Martin DP, Lemey P, Lott M, et al. RDP3: a flexible and fast computer program for analyzing recombination. *Bioinformatics.* 2010; 26:2462–2463. [PubMed: 20798170]
- Posada D, Crandall KA. MODELTEST: testing the model of DNA substitution. *Bioinformatics.* 1998; 14:817–818. [PubMed: 9918953]
- Pybus OG, Barnes E, Taggart R, et al. Genetic history of hepatitis C virus in East Asia. *J Virol.* 2009; 83:1071–1082. [PubMed: 18971279]
- Pybus OG, Markov PV, Wu A, et al. Investigating the endemic transmission of the hepatitis C virus. *Int J Parasitol.* 2007; 37:839–849. [PubMed: 17521655]
- Simmonds P. The origin and evolution of hepatitis viruses in humans. *J Gen Virol.* 2001; 82:693–712. [PubMed: 11257174]
- Simmonds P, Bukh J, Combet C, et al. Consensus proposals for a unified system of nomenclature of hepatitis C virus genotypes. *Hepatology.* 2005; 42:962–973. [PubMed: 16149085]
- Tippmann HF. Analysis for free: comparing programs for sequence analysis. *Brief Bioinform.* 2004; 5:82–87. [PubMed: 15153308]
- Wang Y, Xia X, Li C, et al. A new HCV genotype 6 subtype designated 6v was confirmed with three complete genome sequences. *J Clin Virol.* 2009; 44:195–199. [PubMed: 19179105]
- Wang H, Barnes E, Yuan M, et al. Eight novel hepatitis C virus genomes reveal the changing taxonomic structure of genotype 6. *J Gen Virol.* 2013; 94:76–80. [PubMed: 23015745]
- Xia X, Lu L, Tee KK, et al. The unique HCV genotype distribution and the discovery of a novel subtype 6u among IDUs co-infected with HIV-1 in Yunnan, China. *J Med Virol.* 2008a; 80:1142–1152. [PubMed: 18461611]
- Xia X, Zhao W, Tee KK, et al. Complete genome sequencing and phylogenetic analysis of HCV isolates from China reveals a new subtype, designated 6u. *J Med Virol.* 2008b; 80:1740–1746. [PubMed: 18712831]

- We characterized 16 full-length HCV genomes in China to verify its increased diversity.
- They represent subtypes 1a, 1b, 1c, 2a, 2b, 2f, 3a, 3b, 4d, 5a, 6a, and two novel genotype 6 variants.
- GZ51969 and GZ52540 represent two unique 1b lineages that are prevalent nationwide.
- ZS542 and GZ98799 are isolates of subtype 2f confirming the designation of this subtype for the first time.
- ZS542, ZS537, and ZS631 are isolates of subtypes 1c, 4d, and 5a that are first detected in the country.

Figure 1.

A maximum likelihood tree estimated from the full-length genome sequences of genotypes 1, 3, 4, and 5. Each of the sequences is indicated with a subtype name followed by an isolate ID, and those sequences that were determined in the present study are labeled with a black circle. The bootstrap supports are shown in italics. The scale bar indicates a genetic distance of 0.10 nucleotide substitutions per site.

Figure 2.

A maximum likelihood tree estimated from the full-length genome sequences of genotype 2. Except for the six triangles, which indicate six unclassified isolates, all of the other indications are the same as in Figure 1.

Figure 3.

A maximum likelihood tree estimated from the full-length genome sequences of genotype 6. Except for the H77 strain, which was used as an outlier group, all of the other indications are the same as in Figure 1.

Table 1
Patients' information, amplification strategy, and genomic organization of the 16 isolates

ID	Gender	Age	Date	Genotype	Amplicons	Full	ORF	5'UTR	Core	E1	E2	P7	NS2	NS3	NS4A	NS4B	NS5A	NS5B	3'UTR	Poly T or 3'
ZS30/1a	M	59	2009	1a	15	9,488	9,036	341	573	576	1,089	189	651	1,893	162	783	1,344	1,776	111	T
GZ51969/1b/PYB			2002	1b	13	9,628	9,033	341	573	576	1,089	189	651	1,893	162	783	1,341	1,776	254	X
GZ52540/1b/PYA			2002	1b	13	9,632	9,033	341	573	576	1,089	189	651	1,893	162	783	1,341	1,776	258	X
ZS17/1c	M	40	2009	1c	20	9,424	9,036	324	573	576	1,089	189	651	1,893	162	783	1,344	1,776	64	
ZS623/2a	M	46	2011	2a	18	9,475	9,102	340	573	576	1,101	189	651	1,893	162	783	1,398	1,776	33	
ZS260/2b	M	50	2010	2b	21	9,441	9,102	339	573	576	1,101	189	651	1,893	162	783	1,398	1,776	0	
ZS542/2f	M	73	2010	2f	21	9,472	9,102	340	573	576	1,101	189	651	1,893	162	783	1,398	1,776	30	
GZ98799/2f			2002	2f	28	9,470	9,102	340	573	576	1,101	189	651	1,893	162	783	1,398	1,776	28	
ZS633/3a	M	36	2011	3a	17	9,477	9,066	339	573	576	1,107	189	651	1,893	162	783	1,356	1,776	72	T
ZS650/3b	M	40	2011	3b	18	9,434	9,069	339	573	576	1,110	189	651	1,893	162	783	1,356	1,776	26	
ZS537/4d	M	62	2010	4d	19	9,427	9,021	341	573	576	1,086	189	651	1,893	162	783	1,332	1,776	65	T
ZS631/5a	M	52	2011	5a	16	9,414	9,045	339	573	576	1,092	189	651	1,893	162	783	1,350	1,776	30	
ZS221/6a	F	20	2009	6a	19	9,430	9,060	298	573	576	1,104	189	651	1,893	162	783	1,353	1,776	72	T
ZS674/6a	F	29	2011	6a	19	9,429	9,060	298	573	576	1,104	189	651	1,893	162	783	1,353	1,776	71	T
ZS96/6?	F	63	2009	6?	25	9,488	9,075	338	573	576	1,113	189	651	1,893	162	783	1,359	1,776	75	T
ZS202/6?	M	58	2009	6?	20	9,447	9,054	338	573	576	1,098	189	651	1,893	162	783	1,353	1,776	55	T