Corrections

MICROBIOLOGY

The editors note that on September 10, 2013, PNAS received a letter to the editor from Naccache et al. (1), which raised concerns about a possible contaminant affecting the results of this paper. The authors were asked to provide a response to these concerns, but while we were awaiting their comments, the Office of Research Integrity (ORI) published a case summary (2) for Baoyan Xu on December 30, 2013. Based on a limited admission by Xu, ORI found that Baoyan Xu had engaged in research misconduct. ORI recommended a correction to Fig. 6 and Fig. S4. After evaluation of the authors’ response (3) to Naccache et al. and the detailed information supplied below from the authors, the Editorial Board approved the following correction:

The authors wish to note,” Correspondence from Stephen Korsman, Medical Virologist at Groote Schuur National Health Laboratory Service, South Africa, suggested that some immunoblots in the above article contained multiple copied images, in which images were reproduced for several different patients. Using graphics-editing software (varying contrast and lighting of individual lanes of the blots), we confirmed that the images had been extensively manipulated. This finding led to a formal complaint to and subsequent investigation by the Department of Health and Human Services ORI. ORI has formally concluded that one author engaged in research misconduct (2). Misconduct appeared to be restricted to assembly of the final, complex figures.

“We were able to retrieve and reconstruct digital images from the original experiments; only immunoblots from patients with seronegative hepatitis, in Fig. 6B and Fig. S4,A, had been manipulated. We further repeated immunoblotting experiments, using existing samples, and we used these data to replace both Fig. 6 and Fig. S4. This correction does not affect the original findings of the serology studies. We have evaluated additional healthy controls from different ethnicities, which we add to the Supporting Information as Fig. S4C.

“We reproduced the experiments using sera samples from 77 Chinese hepatitis patients by immunoblotting as described in our paper. Fifteen patient sera samples that were used in the original study were unavailable. In addition, we tested sera samples from 28 healthy controls with origins from the United States, France, Korea, Brazil, China, Vietnam, Japan, and Puerto Rico. By immunoblot, 86% (66/77) of seronegative hepatitis patients were positive for NIH-CQV IgG, and 16% (12/77) were positive for IgM (Fig. 6B and Fig. S4A). Of healthy controls from different ethnicities, 78% (13/28) were positive for IgG (Fig. S4C). These data confirm the presence of an antibody response against recombinant capsid protein of NIH-CQV in the cohort of Chinese seronegative hepatitis patients, and also in healthy individuals. Reactivity between the polyhistidine-tag and samples also was evaluated and excluded. However, whether the antibody response against recombinant NIH-CQV viral capsid protein detected by immunoblot was induced by infection with NIH-CQV or derives from cross-reactivity with an epitope present on another microorganism remains to be clarified. Although we excluded cross-reactivity between NIH-CQV and other common human parvoviruses, additional confirmatory experiments based on serologic reactivity to multiple, nonoverlapping fragments derived from NIH-CQV capsid protein are in progress and the results will be reported.”

The corrected figure and its corrected legend appear below.

Fig. 6. Immunoblot of seronegative hepatitis patient using recombinant capsid protein (rCP) of NIH-CQV. (A) Specificity test for the rCP of NIH-CQV. The cell lysates derived from the cells transfected with the plasmids that expressed capsid proteins of AAV2, Parv4, HBoV, and B19V and purified rCP were subjected to SDS/PAGE, and then transferred to a nitrocellulose membrane. After blocking with 5% (wt/vol) nonfat milk for 2 h, the membrane was incubated with respective antisera at 1:1,000 dilution. (B and C) Detection of specific antibodies against the rCP of NIH-CQV. Samples subjected to SDS/PAGE consisted of 25 ng of affinity-purified rCP. These proteins were transferred to a nitrocellulose membrane and incubated with a 1:1,000 dilution of individual serum. Above the lanes are the identification numbers of individual patients and healthy controls. Only representative results are shown; see Fig. S4 A and B for complete data. Mouse monoclonal antibody against the polyhistidine tag was used as a positive control. The numbers on the left indicate molecular masses in kilodaltons based on the PageRuler Prestained Protein Ladder (Fermentas).
**SYSTEMS BIOLOGY**


The authors note “that a mistake was made generating the Epl1 mutant EPCa-Q shown in Fig. 3 B and C and Fig. S4. Instead of mutating K39, 345, 376 and 379 to Q (glutamine, Gln) these sites were mutated to E (glutamic acid, Glu). Through gene synthesis, we have generated the correct K39L, 345Q, 376Q, and 379Q mutant and assessed its impact on NuA4. As anticipated, detectable acetylation signal on Epl1 mutants EPCa-R and EPCa-Q was decreased (corrected Fig. 3B). We next sought to determine the function of Epl1 acetylation within the EPCa domain. Although the EPCa-Q and EPCa-R mutants do not impact coimmunopurification with Eaf5-TAP, global histone H4 acetylation was reproducibly decreased in esal-L254P and EPCa-R (corrected Fig. 3B, lanes 2 and 3), in contrast EPCa-Q displayed global histone H4 acetylation similar to wild type Epl1 strains (corrected Fig. 3B, lanes 1 and 4). In agreement with this finding, the strain expressing EPCa-Q does not display temperature sensitivity (corrected Fig. S4). In conclusion, while mutating K39, 345, 376, and 379 to acidic glutamic acid has negative consequences for NuA4 catalytic activity (original Fig. 3), mutating these sites to glutamine does not. Therefore, as reported for other MYST KATs (1), autoacetylation does contribute to NuA4 catalytic activity. The authors apologize for this error.”

Please see the table below for new strains.”

**Strain list**

<table>
<thead>
<tr>
<th>Strain</th>
<th>Genotype Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>YKB3402</td>
<td>MATa his3 leu2 lys2 ura3 trp1 epil1::kanMX EAF5-TAP::HIS3 [epil1-K39Q-K345Q-K376Q-K379Q-3HA::LEU2] Present study</td>
</tr>
<tr>
<td>YKB3579</td>
<td>MATa ura3-52 lys2-801 ade2-101 trp1-alpha63 his3-J320 leu2-117 epil1::KAN [epil1-K39Q-K345Q-K376Q-K379Q-3HA::LEU2] Present study</td>
</tr>
</tbody>
</table>

As a result of these errors, the authors note that the following section should be added to the Materials and Methods.

**Cloning and Mutagenesis of Enhancer of Polycomb Like 1.** Acetylation point mutants mimicking acetylated (K-to-Q) state for each acetylated lysine were generated by synthesizing a custom-designed DNA fragment (Genescript) of 857 base pairs carrying the mutation K345Q, K376Q, and K379Q and corresponding to nucleotides 600–1456 of EPL1. Following double digestion with BspE1/Nru1 and gel purification (Qiagen; 28704), this mutated DNA fragment was ligated into pRS415-EPL1-HA3 (pKB41) that was also digested with BspE1/Nru1, subjected to calf intestinal phosphatase treatment (New England Biolabs; M0290) and gel purified. Ligation was carried out overnight at 16°C with T4 DNA ligase (New England Biolabs; M0202S), and then transformed into competent Escherichia coli DH5α cells. K39Q was then generated by using a site-directed mutagenesis kit (Stratagene; 200528). The successful introduction of all point mutations pRS415-epil1-EPCa-Q-HA3 (pKB264) was confirmed by sequencing and expression confirmed by Western blot analysis.”


The authors also note that Fig. 3 and its corresponding legend appeared incorrectly. The corrected figure and its corrected legend appear below.

---

**Fig. 3.** NuA4 subunits are acetylated in vivo and in vitro. (A) Radioactive KAT assay to assess NuA4 auto-acetylation in vitro. [3H]-acetyl CoA was added directly to bead matrix of a stringently immunopurified NuA4 preparation [Esa1-TAP (YKB440)] versus an untagged control (YPH499) immunopurification. Histone proteins were added to the reaction to serve as a positive control for acetylation. The reactions were separated on a gradient gel, Coomassie-stained to visualize proteins (Right), treated for fluorography, and finally exposed to film for 2 wk (Left). Proteins are identified on the right side and protein size is indicated on the left (kDa). (B) Epl1 acetylation is dependent on Esa1 in vivo. NuA4 was purified from cells grown at 25°C through Eaf5-TAP. Epl1 was expressed as a C-terminal HA fusion protein in epl1Δ background strain in the presence (lane 2, YKB2876) or absence (lane 1, YKB2862) of the acetyltransferase-deficient esal-L254P allele, or as a lysine-to-arginine or -glutamine multipoint mutant [EPCa-R; EPCa-Q (K39,345,376,379R or Q) YKB2781 and YKB3402, lanes 3 and 4, respectively]. Immunopurified (IP) products and whole cell extract (WCE) were separated by SDS/PAGE (7.5%) and subjected to Western blot using the indicated antibodies: antiacetyl lysine (α-AcK), antihistone H4 acetyl lysine (α-AcK-H4), anti-γ-glutarylaldehyde 5-phosphate dehydrogenase (α-G6PDH). Image is a representative of three independent replicates. (C) Esal1-TAP-EPCa-R reduces NuA4 in vitro KAT activity. NuA4 was purified through Eaf5-TAP (lanes 2–4) relative to an untagged control sample (lane 1, YPH499). Epl1 was expressed from its endogenous locus (lane 2, YKB1042) or as a C-terminal HA fusion protein of wild type EPL1 (lane 3, YKB2862) or epil1-EPCa-R (lane 4, YKB2781). KAT assays were performed using NuA4 preparations equalized for Esa1 and nucleosome purified from HeLa cells. Error bars represent SD from duplicate reactions. NuA4 complexes used in assay were separated by SDS/PAGE and subjected to Western blot using the indicated antibody anti-Esa1 (α-Esa1) (Lower).

www.pnas.org/cgi/doi/10.1073/pnas.1401790111
Hybrid DNA virus in Chinese patients with seronegative hepatitis discovered by deep sequencing

Baoyan Xu,a,b,1 Ning Zhi,a,b,2 Gangqing Hu,a,c,1 Zhihong Wan,a Xiaobin Zheng,d Xiaohong Liu,a Susan Wong,a Sachiko Kajigaya,a Keji Zhao,c,3 Qing Mao,b,2 and Neal S. Youngb,3

*Hematology Branch and bSystems Biology Center, National Heart, Lung, and Blood Institute, Bethesda, MD 20892; bInstitute of Infectious Disease, Southwest Hospital, Third Military Medical University, Chongqing 400038, China; cDepartment of Embryology, Carnegie Institution for Science, Baltimore, MD 21218

Edited* by Harvey Alter, National Institutes of Health, Bethesda, MD, and approved March 19, 2013 (received for review March 4, 2013)

Seronegative hepatitis—non-A, non-B, non-C, non-D, non-E hepatitis—is poorly characterized but strongly associated with serious complications. We collected 92 sera specimens from patients with non-A–E hepatitis in Chongqing, China between 1999 and 2007. Ten sera pools were screened by Solexa deep sequencing. We discovered a 3,780-bp contig present in all 10 pools that yielded BLASTx E scores of 7e-05-0.008 against paroviruses. The complete sequence of the in silico-assembled 3,780-bp contig was confirmed by gene amplification of overlapping regions over almost the entire genome, and the virus was provisionally designated NIH-CQV. Further analysis revealed that the contig was composed of two major ORFs. By protein BLAST, ORF1 and ORF2 were most homologous to the replication-associated protein of bat circovirus and the capsid protein of porcine parovirus, respectively. Phylogenetic analysis indicated that NIH-CQV is located at the interface of Paroviridae and Circoviridae. Prevalence of NIH-CQV in patients was determined by quantitative PCR. Sixty-three of 90 patient samples (70%) were positive, but all those from 45 healthy controls were negative. Average virus titer in the patient specimens was 1.05 e4 copies/mL. Specific antibodies against NIH-CQV were sought by immunoblotting. Eighty-four percent of patients were positive for IgG, and 31% were positive for IgM; in contrast, 78% of healthy controls were positive for IgG, and all were negative for IgM. Although more work is needed to establish the etiological role of NIH-CQV in human disease, our data indicate that a parovirus-like virus is highly prevalent in a cohort of patients with non-A–E hepatitis.

Most viral hepatitis is secondary to infection by known hepatotropic viruses: hepatitis A virus (HAV) (1), hepatitis B virus (HBV) (2), hepatitis C virus (HCV) (3), hepatitis delta virus (HDV) (4), and hepatitis E virus (HEV) (5). Other hepatitis-associated viruses, including CMV (6), Epstein–Barr virus (EBV) (7), herpes simplex virus 1 and 2 (8), varicella-zoster virus, human herpesvirus 6 (9), human parvovirus B19V (10), and adenoviruses (11), may cause liver injury, ranging from mild and transient elevation of aminotransferases to acute hepatitis and occasionally acute liver failure. Despite the number of hepatitis viruses and hepatitis-associated viruses, etiology cannot be determined in 10–20% of the cases of acute hepatitis (12), in 30% of cases of cryptogenic chronic liver diseases (13), in hepatitis-associated aplastic anemia (14), and in a large proportion of cases of acute liver failure (15). Patients with acute seronegative hepatitis have fewer parentral risk factors, more severe bilirubin and transaminase elevations, and worse hepatocellular synthetic function than do patients with hepatitis C (16). Moreover, seronegative hepatitis is more strongly associated with serious complications, especially bone marrow failure (14) and childhood fulminant hepatitis (17).

Next-generation sequencing (NGS) technology has had an increasing impact on biological research and clinical diagnosis, because it provides high-resolution genome-scale data rapidly and reliably. The first application of NGS for pathogen discovery in a patient who died of a febrile illness after solid-organ transplantation led to the identification of an arenavirus (18). NGS-based platforms, such as 454 and Solexa deep sequencing, have been applied successfully to the investigation of emerging human pathogens, resulting in major discoveries of new human viruses (19–21).

In the present study, we established an experimental and analytic procedure for identifying virus identification in human sera specimens based on Solexa deep sequencing. We applied the method to screen for potential viral infection in human sera specimens and discovered a human virus in Chinese patients who had seronegative hepatitis. From its genome organization and phylogeny, the virus is unusual, because it is evolutionarily at the interface of the parovirus and circovirus families, perhaps having emerged from inter familial recombination. The virus was highly prevalent in a patient population, and active infection in seronegative hepatitis cases was suggested by the presence of viral DNA in the blood and serology. However, more work is needed to establish a firm disease association.

Results

Detection of a Virus in Patients with Seronegative Hepatitis. Ten pools were made of sera from 92 patients with non-A–E hepatitis and were screened for potential pathogens by Solexa deep sequencing. After filter sterilization, the samples were digested with DNase and RNase to eliminate host nucleotide contamination, and the remaining nucleic acids were extracted using carrier RNA [synthetic poly(A)]. cDNA was synthesized from extracted viral nucleic acids with non-poly(A) random hexamers, which were designed specifically to block reverse transcription of the carrier RNA. With non-poly(A) random hexamers, redundant sequence tags derived from the carrier RNA were reduced to a level of <0.001%. High capacity enabled multiplex sequencing of more than 70 individual samples in a single Solexa run. Short DNA sequences were analyzed and filtered at the nucleotide level to exclude potential DNA contamination from human cells, mouse cells, and known bacteria. For detection of known viruses, non-redundant reads were mapped against virus genomes from the National Center for Biotechnology Information (NCBI) Reference Database (accession no. KC617868).


The authors declare no conflict of interest.

Data deposition: The sequence reported in this paper has been deposited in the GenBank database (accession no. KCC17868).

1B.X., N.Z., and G.H. contributed equally to this work.

2To whom correspondence may be addressed. E-mail: zhin@nhlbi.nih.gov or qingmao@yahoo.com.

3K.Z. and N.S.Y. contributed equally to this work.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1303744110/-/DCSupplemental.
Sequence (RefSeq) database using the short sequence alignment tool Bowtie. HAV and HCV sequences were detected in samples derived from sera pools 2 and 4, respectively. Individual samples from these two pools then were sequenced to clarify the source of the hepatitis viruses, and high copy numbers of HAV or HCV genomes were detected in one patient in each pool (these two samples were excluded from further analysis of seronegative hepatitis cases). Furthermore, sequences from Adenoviridae, Herpesviridae, Paroviridae, and Anelloviridae were also commonly detected. For virus discovery, filtered short sequences were subjected to de novo assembly. Contigs from different samples were clustered based on nucleotide-level similarities. The resulting contigs were further compared with the NCBI nonredundant (NR) protein database using tBLASTx. A contig was predicted as viral origin if the best tBLASTx hit was from the kingdom of viruses, and a cluster was defined as a candidate if it contained a contig with viral origin. Contigs for each cluster were subjected to multiple alignments to define a consensus contig for later data analysis and experimental validation.

By applying this pipeline to the 10 pooled samples, we identified a 3,780-bp contig that yielded BLASTX E scores of 7e-05-0.008 against parvoviruses (GenBank accession no.: KC617868).

Characterization of a Human Virus Genome. The 3,780-bp contig was composed of three major ORFs (Fig. 1 and Fig. S1). Protein Blast (BLASTp) searches showed that ORF1 encoded a 45-kDa protein, which contained a domain (GxxxxGK[T/S]) (22) for a phosphate-binding loop (P-loop) that is conserved among different DNA viruses (Table S1); ORF1 was homologous to the replication-associated protein (Rep) of bat circovirus with E values of 7e-04. ORF2 encoded a 55-kDa protein which was homologous to Parovirus coat protein 1 (VP1) of porcine parvovirus (PPV) and goose parvovirus, with E scores of 2e-05. Because of low homology, only the first 87 amino acid residues encoded by ORF2 could be aligned reliably to the first 100 amino acids of VP1 of PPV. The first 100 amino acids of the N-terminal region of the PPV VP1 region constituted an active phospholipase A2 (PLA2), which is highly conserved among the members of the Paroviridae and is essential for parvovirus infection (23). The putative capsid protein (CP) encoded by ORF2 included a PLA2-like motif that is highly conserved among the members of the Paroviridae (Table S2). ORF3 was located on the left side of the genome on the viral negative strand and encoded a 15-kDa protein. There was no homology at the nucleic acid level between the 3,780-bp contig and known viruses in GenBank.

Sequence analysis also revealed inverted terminal repeats (ITR) at both ends of the genome (Fig. S1). The ITR was 156 nucleotides long. The first 75 nucleotides complemented nucleotides 82–156, allowing the formation of a stem of a hairpin, whereas nucleotides 78–81 were mismatched and could form a loop. By sequence analysis of the intergenic region between ORF1 and ORF3, we found multiple direct repeats and TATA boxes (Fig. S1). In silico promoter prediction suggested that consensus sequences for a bidirectional euakaryote promoter were present in this region (www.fruitfly.org/seq_tools/promoter.html) and that the NIH-CQV might have an ambisense genome.

To confirm the sequence of the de novo-assembled 3,780-bp contig, eight pairs of PCR primers (Table S3) were designed, based on the sequence predicted and were used to amplify overlapping DNA fragments from the pooled patient samples. Eight amplified DNA fragments were of the expected lengths (Fig. S2), and their sequences could be assembled into a 3,629-bp contig. Sequencing analysis revealed that the 3,629-bp contig, composed of the entire coding region plus half of the ITRs from both 5’ and 3’ termini, aligned exactly with that of the de novo-assembled 3,780-bp contig, indicating that the contig contained the nearly complete virus genome, was indeed present in the patient samples, and was not an artifact generated by assembly. The virus was provisionally designated NIH-CQV virus (NIH-CQV), because the samples were collected in a hospital in Chongqing and the laboratory experiments were conducted at the National Institutes of Health (NIH).

Because ORF1 of NIH-CQV encodes a putativeRep protein homologous to circovirus Rep, we speculated that the virus might have a circular form of its genome. To test this hypothesis, we performed inverse PCR with a primer pair (Table S3) oriented outwardly with respect to each other. Amplification with the inverted-primer pair generated an amplicon of 116 bp from two of three patient samples tested. Sequencing and alignment of the inverse PCR product showed a junction region between the 5’ and 3’ termini in a head-to-tail orientation. In comparison with the linear genome of NIH-CQV, a 419-bp region composed of both 5’ and 3’ ITRs present in the linear virus genome was absent in the inverse PCR product (Fig. S3), indicating a closed circular form of the virus genome of 3,361 bp present in the patient samples. Circularization of the viral genome resulted in an extension of the ORF3 at the left end of the genome, from 369 bp to 435 bp, with predicted encoding of a 17-kDa protein. The circular form of NIH-CQV was designated NIH-CQV-Co.

Phylogenetic and Evolutionary Analyses of NIH-CQV. Because ORF1 and ORF2 of NIH-CQV were homologous to different viral families, phylogenetic analysis of the two ORFs was performed separately by comparison with members of Circoviridae and Paroviridae. Although BLASTp searches showed that the amino acid sequences encoded by ORF1 were homologous to Rep of bat circovirus, by gene tree analysis NIH-CQV was not closely related to any known circoviruses (Fig. 2A). For ORF2, the amino acid similarity between NIH-CQV and known parvoviruses was below 20%. Because of low homology, only the first 87 amino acid
residues encoded by ORF2 could be aligned reliably with the VP1 of porcine parvovirus and goose parvovirus, and in this region amino acid identity was ~31%. NIH-CQV was not closely related to any known parvoviruses and thus appeared to be deeply rooted by lineage between human and animal parvoviruses and parvoviruses identified in arthropods (Fig. 2B). To confirm further the phylogenetic location of NIH-CQV, whole-proteome phylogeny analysis of NIH-CQV with 14 circoviruses and 21 prototypes of parvoviruses was conducted using a dynamic language model (24), which correctly divided the 35 selected viruses into two families, Parvoviridae and Circoviridae, and placed NIH-CQV at the interface of Parvoviridae and Circoviridae (Fig. 3).

Deep-sequencing data showed that NIH-CQV exhibited remarkable intrapatient genetic heterogeneity, suggesting the presence of closely related variants or quasispecies in the patients infected with NIH-CQV. To assess the dynamics of circulating quasispecies, we measured the ratios of nonsynonymous (Ka) versus synonymous (Ks) substitutions across the entire rep and cp of NIH-CQV from patients with acute (n = 13) or chronic (n = 9) non-A–E hepatitis. A higher value of Ka/Ks reflects selection favoring amino acid mutation (positive selection), whereas lower values are consistent with selection against nonsynonymous variants (negative selection). Positive selection results in the dominance of virus variants with a survival advantage, whereas negative selection acts to eliminate variants with a diminished capacity to complete the virus-replication cycle. The average Ka/Ks ratio of the cp of the virus was greater in patients with chronic hepatitis than in patients with acute hepatitis (0.98 ± 0.27 versus 0.71 ± 0.17; P < 0.01) (Fig. 4A). There were three regions in cp, ranging from nucleotides 2,360–2,448, 2,562–2,732, and 2,969–3,036 of the NIH-CQV genome, in which Ka/Ks ratios were greater than 1 and were significantly higher than in other regions (P < 0.05) (Fig. 4B), suggesting positive selection acting on specific regions of the cp gene. In contrast, although the average Ka/Ks ratio of the rep from patients with chronic hepatitis was greater than that for patients with acute hepatitis [0.47 ± 0.08 versus 0.39 ± 0.08 (P < 0.05)] (Fig. 4C), the overall Ka/Ks ratio was much lower (0.42), and there was no evidence for positive selection of this gene (Fig. 4D).

Prevalence of the NIH-CQV Virus in Clinical Specimens. We designed a quantitative PCR (qPCR) assay for the rep region of NIH-CQV and generated a standard curve using a synthetic DNA rep fragment. Sensitivity of detection was <10 copies. We conducted screening of the 90 sera specimens from the patients with seronegative hepatitis and 45 sera specimens from healthy blood donors. Seventy percent (63/90) of the patient samples were positive, but all the samples from healthy controls were negative. The average virus titer in the patient specimens was 1.05 e4 copies/μL, and the highest was 3.1 e4 copies/μL (Fig. 5). No significant difference in NIH-CQV DNA levels was observed between patients with acute or chronic hepatitis (Table S4). The two patients with cryptic HAV and HCV infection were negative for NIH-CQV by qPCR.

Detection of Specific Antibody Against NIH-CQV in Patients with Non-A–E Hepatitis. To investigate the immune response against NIH-CQV, a synthetic DNA fragment encoding a 184-amino acid polypeptide of the C-terminal portion of CP was cloned into an expression vector, and the affinity-purified recombinant CP (rCP) was used for serological studies. Cross-reactivities between NIH-CQV and other human parvoviruses, including B19V, human parvovirus 4 (Parv4), human bocavirus (HBoV), and adenov-associated virus 2 (AAV2), were tested by immunoblotting. No cross-reactivity was seen between CP of NIH-CQV and other major human parvoviruses (Fig. 6A). Immunoreactivity of rCP with a total of 135 human sera specimens was examined. By immunoblotting, 84% (76/90) of the patients with seronegative hepatitis were positive for NIH-CQV IgG, and 31% (28/90) were positive for IgM, suggestive of recent infection. In contrast, 78% (35/45) of healthy controls were positive for IgG, but all were negative for IgM, consistent with past NIH-CQV exposure (Fig. 6B, Fig. S4A and B, and Table 1). There was a significant difference in NIH-CQV–specific IgM, but not in IgG, between the patients with acute and chronic hepatitis (Table S4).

Discussion

We report the identification of a virus, NIH-CQV, in patients with non-A–E hepatitis. Phylogenetically NIH-CQV appears to lie at the interface of parvoviruses and circoviruses. The genome of Fig. 3. Whole-proteome tree of NIH-CQV and members of Paroviridae and Circoviridae. A total of 18 circoviruses and 28 parvoviruses are included in the analysis. The tree was constructed by neighbor-joining based on protein sequences using dynamic language method for K = 4. The different viral genera are color-coded to the branches of the tree based on the taxonomic partitions. Detailed information regarding abbreviations and accession numbers are available in Dataset S18.

Fig. 4. Quasispecies of NIH-CQV in patients with acute and chronic hepatitis. (A and C) Scatterplot presentations of Ks/Ka ratios of rep and cp genes of NIH-CQV from patients with acute or chronic hepatitis. The average values of Ks/Ka ratios of cap or rep of NIH-CQV from 22 individual patients (acute: n = 13; chronic: n = 9) are indicated. (B and D) Distribution of Ks/Ka ratios throughout the cp and rep. The y-axis is the averaged values of Ka or Ks within a sliding window of 100 amino acids. The average values of Ks/Ka across the whole gene region of NIH-CQV from the 22 patients are indicated. Genomic regions with significant high Ks/Ka (P < 0.05; binomial test) are shown in gray.

Xu et al.
NIH-CQV has no homology to any known viruses in GenBank. Although comparative analysis revealed that the Rep and CP proteins of NIH-CQV have limited homologies with circoviruses and parvoviruses, respectively, the overall genome organization of NIH-CQV has the basic characteristics of viruses in Paroviridae. NIH-CQV has a small, compact linear DNA genome with two tandemly arranged major ORFs encoding Rep and CP proteins, respectively, and a pair of ITRs at the left and right ends of the genome. In silico promoter prediction suggested that transcription of the rep and cp was regulated by a single promoter upstream of the rep, because there was no apparent promoter sequence immediately upstream of cp. A pre-mRNA encoded by a single promoter that is processed by alternative splicing and alternative polyadenylation to generate multiple mRNAs had been reported for other parvoviruses, such as B19V (25) and Aleutian mink disease virus (26). By amino acid sequence analysis, CP protein of NIH-CQV is minimally homologous to porcine and goose parvovirus VP1 (27). The homologous region is mainly in the N terminus of CP and contains the conserved PLA2-like motif, which is present in the N-terminal extension of the VP1 unique region of members of Paroviridae (23). The phylogenetic tree constructed using amino acid sequences of CP of NIH-CQV and VP1 of other representative viruses in Paroviridae showed that NIH-CQV represents a deeply rooted lineage between two groups: (i) human and animal parvoviruses and (ii) insect parvoviruses. Thus, NIH-CQV appears to be a parvovirus-like virus. However, NIH-CQV also has features shared by members of the Circoviridae family (28). First, a BLASTp search revealed that the Rep of NIH-CQV is more related to circoviral Rep than to those of other viruses. Second, NIH-CQV appears to have an ambisense genome because there are bidirectional putative promoters in the intergenic region between the rep and 15-kDa protein and because rep and 15-kDa protein arranged head-to-head flank the intergenic region, which is a feature of circoviruses. Finally, sequencing and alignment of the inverse PCR products indicated the existence of the circular genome of NIH-CQV in the clinical samples. The circular form of NIH-CQV genome might represent a subviral DNA or an intermediate structure of viral replication, as reported for other parvoviruses (29). Therefore, NIH-CQV appears to be a “hybrid” virus, perhaps formed by inter familial recombination between a parvovirus and a circovirus. This idea was supported further by whole-proteome phylogenetic analysis, which showed NIH-CQV located between Paroviridae and Circoviridae.

Both recombination and a high mutation rate are often cited as key in evolutionary innovation. For many viruses, recombination...
allows single-step acquisition of multiple genetic changes, and a high mutation rate enhances adaptation to new hosts (30). These processes are critical driving forces in viral evolution and lead to viral emergence, host-switching, and adaptation, which often result in disease outbreaks (31, 32). Recent studies had demonstrated that small eukaryotic ssDNA viruses have high rates of nucleotide substitutions and genetic recombination, which may surpass rates seen in RNA viruses (33, 34). These observations may explain the extensive diversity seen in ssDNA viruses. Parvoviruses (family *Parvoviridae*) and circoviruses (family *Circoviridae*) are both ssDNA viruses that infect a wide variety of animal species, including mammals, birds, and insects. Although horizontal gene transfer and high mutation rates have been documented in these viruses, cross-species viral transfers resulting from interfamilial recombination and leading to a infection in the new host species are not common.

A high rate of viral evolution has been reported for emerging parvoviruses (35). Canine parvovirus (CPV) is one well-characterized example (35). CPV emerged from feline panleukopenia parvovirus or a closely related virus, differing at several amino acid residues. Sequences of the viruses collected at time points before and after host-switching suggest that the emergence of CPV was dependent on a high mutation rate and positive selection of the major capsid gene. Our deep-sequencing data showed that NIH-CQV has a high substitution mutation rate throughout its genome. However, the impact of selective pressures, as measured by the $K_s/K_a$ ratio, appeared to vary depending on location. The $cp$ gene contained three regions in which the $K_s/K_a$ ratio exceeded 1, suggesting positive selection. These substitutions would be presumed to enhance fitness. In contrast, selective pressure on the $rep$ of NIH-CQV appeared to be dominated by purifying selection, because the average $K_s/K_a$ ratio for the $rep$ from 14 patients was only 0.42. The $K_s/K_a$ ratios for the regions encoding the PLA$_2$ motif in the CP protein or P-loop nucleotide triphosphatase (NTPase) in the Rep proteins were lower than in other regions, as is consistent with strong purifying selection and functional importance of this region in viral replication.

NIH-CQV exhibited remarkable genetic heterogeneity within patients, indicating the presence of closely related variants or quasispecies in NIH-CQV–infected individuals. The $K_s/K_a$ ratios of both viral $cp$ and $rep$ were greater in patients classified as having chronic hepatitis than in those with acute hepatitis, suggesting greater variation or more diversity of quasispecies of NIH-CQV over time. For hepatitis viruses C and E, diversification of quasispecies during infection is driven by the interaction between the viruses and host immune response and may be associated with different clinical outcomes (36, 37). With a limited number of serial samples from a single individual, we were unable to conclude whether quasispecies dynamics of NIH-CQV in patients correlated with disease.

Despite technological advances in molecular biology, an etiology cannot be determined in as many as 20% of hepatitis cases. As of yet, no causative agent may be defined in a significant proportion of patients with chronic liver disease and cirrhosis, whose disease therefore is characterized as “cryptogenic.” Additional hepatitis agents have been suspected. Over several decades, several human viruses, such as GB virus C (38), Torque teno virus (39), and SEN virus (40), have been putatively claimed to be hepatitis viruses, but subsequent investigation revealed no conclusive association between these viruses and liver disease. Therefore, the etiology of non-A–E hepatitis remains unresolved. In this study, we applied NGS methodology to virus discovery and indentified a parvovirus-like virus in the sera samples collected from a cohort of patients with non-A–E hepatitis. Our results showed that 28% of patients were positive for both NIH-CQV–specific IgM and viral DNA, suggesting recent infection in patients. Most patients and healthy controls were positive for virus-specific IgG, suggesting that exposure was common in the population studied. The patients who yielded positive specimens suffered from symptoms of hepatitis with a wide range of liver dysfunction, including acute and chronic hepatitis. There was no difference between patients diagnosed with acute versus or chronic hepatitis with respect to detection of NIH-CQV DNA, but viral DNA was not detected in healthy controls. Although these findings are suggestive of an etiologic relationship, our study has limitations. By electron microscopy, we observed virus-like particles in NIH-CQV qPCR-positive sera which codemented with viral DNA, but we were unable to confirm their identity by immunogold labeling with our antisera. In addition, because of the limited availability of samples, we had insufficient material for viral propagation and isolation in cell culture, and efforts are in progress to detect viral proteins in a few liver biopsy tissue blocks. Additionally, with the limited number of samples available, we cannot exclude the possibility of reactivation of a latent virus in the course of liver inflammation.

In general, parvoviruses cause systemic infection, but clinical manifestations of the infection are varied and may depend on specific tissue tropisms as well as on host immunologic competency. For two human pathogenic parvoviruses, B19V (10) and HBoV (41), clinical symptoms range from none to acute systemic manifestations involving skin and joints to subtle persistence of infection; in immunocompromised populations, B19V chronic infection causes pure red cell aplasia. In addition, coinfection of parvoviruses with other viruses, especially related DNA viruses, can lead to synergy and more severe clinical symptoms in the affected host. Because of the complexity of disease associations, more comprehensive epidemiology and in vitro studies are needed to establish a pathogenic role of NIH-CQV in hepatitis.

### Materials and Methods

**Study Subjects.** We studied a total of 92 sera samples from patients with non-A–E hepatitis who were admitted to the Institute of Infectious Disease of Southwest Hospital, Third Military Medical University, China, between 1999 and 2007. According to the criteria of the Chinese Society of Infectious Disease and Parasitology, and the Chinese Society of Hepatology (42), 33 patients were diagnosed with acute hepatitis by clinical and biochemical parameters. Fifty-nine patients had chronic aggressive hepatitis confirmed by biopsy, and 10 of them had cirrhosis; eight outpatients with chronic persistent hepatitis did not have liver biopsies performed. The patients had a median age of 42 y (range: 12–73 y); among of them, 55 males had a median age of 41 y (range: 12–73 y), and 37 females had median age of 43 y (range: 15–70 y). On admission, mean liver function tests were as follows: total bilirubin 174 ± 130 μmol/L (range: 14–735 μmol/L), alanine aminotransferase 421 ± 623 IU/L (range: 5–3,150 IU/L). Sero- logical tests for anti-HAV IgM (HAV IgM, ELISA Kit; Wantai Biological Pharmacy),

### Table 1. Results of real-time PCR and immunoblot analysis

<table>
<thead>
<tr>
<th></th>
<th>Seronegative hepatitis patients</th>
<th>Healthy controls</th>
</tr>
</thead>
<tbody>
<tr>
<td>qPCR</td>
<td>IgM +</td>
<td>IgG +</td>
</tr>
<tr>
<td></td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>+</td>
<td>27.8% (25/90)</td>
<td>42.2% (38/90)</td>
</tr>
<tr>
<td>-</td>
<td>3.3% (3/90)</td>
<td>26.7% (24/90)</td>
</tr>
<tr>
<td>Total</td>
<td>31.1% (28/90)</td>
<td>68.9% (62/90)</td>
</tr>
</tbody>
</table>

Xu et al.
HBV markers (HBsAg and HBcAg; Roche Diagnostic Systems) anti-HCV antibodies (HCV ELISA Kit; Wantai Biological Pharmacy), anti-HDV IgM (HDV IgM ELISA Kit; Wantai Biological Pharmacy), anti-HIV IgM (HIV IgM ELISA Kit; Wantai Biological Pharmacy), HCV RNA tests (HCV Quantitative RT-PCR Kit; Kehua Bio-engineering), anti-HIV-1 and 2 antibodies (HIV ELISA Kit; Wantai Biological Pharmacy), HIV viral load test (HIV Quantitative RT-PCR Kit; Roche Diagnostic Systems), and serological tests for anti-CMV IgM and (CMV IgM ELISA Kit; Wantai Biological Pharmacy) and anti-EBV IgM (EBV IgM Kit; YHLO Biotech) were all negative. Additional tests for antinuclear antibody, rheumatoid factor, and antimitochondrial antibody (MicroELISA Test Systems for Autoimmunity; EUROIMMUN), as well as blood cultures, urine cultures, and throat swabs for bacteria were negative. The research protocol was approved by the Human Bioethics Committee of the Third Military Medical University, and all participants provided written informed consent.

Solexa Deep Sequencing and Phylogenetic and Evolutionary Analysis. Details of deep sequencing and phylogenetic and evolutionary analysis are provided in SI Material and Methods.

Overlapping PCR and Reverse PCR. To verify the sequence of the viral genome assembled from the Solexa data, eight sets of overlapping primer pairs were designed (Table S3). Viral DNA was extracted from sera as described in the Solexa deep sequencing. Extracted DNA (5 μL) was used as template for the PCR. The 50-μL reaction mix consisted of 1× Platinum Taq PCR Buffer (Invitrogen), 0.2 μM of each dNTP at 0.2 mM, and 25 pmol each of the primers. After 2 min at 94 °C, 35 cycles of amplification (94 °C for 1 min, 54 °C for 1 min, and 72 °C for 1 min) were performed. To detect circularized viral DNA, inverse PCR with a primer pair (Table S3) that oriented outwardly with respect to each other was used for amplification. PCR was performed as described above, and amplified products were visualized on an agarose gel. All PCR products were sequenced.

Diagnostic qPCR. For quantitative PCR, DNA was extracted by the QIAamp MinElute Virus Kit (Qiagen), and 5 μL of the resulting DNA was used for analysis with the NIH-CQV rep primers set (NIH-CQV_Rep-1295F and NIH-CQV_Rep-1470R), and NIH-CQV_Rep probe was used for qPCR (Table S3). All reactions were performed using the Chromo4 real-time detector (Bio-Rad). The reaction started with activation of the polymerase at 95 °C for 15 min, followed by 45 cycles of 15 s at 94 °C and 1 min at 60 °C. The quantitation of amplicon was performed by interpolation with the standard curve to the synthesized rep gene (ORF1) with serial dilutions.

Immunoblotting. To investigate the immune response against NIH-CQV, a synthetic DNA fragment encoding a 184-amino acid polypeptide of the C-terminal portion of CP was cloned into an expression vector, and the affinity-purified CP was used for immunoblotting. Cross-reactions between NIH-CQV and other human paroviruses, including B19V, Parv4, HBoV, and AA2V, were tested by immunoblotting. Details are provided in SI Material and Methods.

ACKNOWLEDGMENTS. We thank Dr. Jun Zhu at the DNA Sequencing and Genomics Core of the National Heart, Lung, and Blood Institute (NHLBI) for assistance in deep-sequencing and data analysis; Dr. Amit Kapoor of Columbia University for technical advice; and Genoveffa Franchini of the National Cancer Institute, Leonid Margolis of the National Institute of Child Health and Development, Cynthia Dunbar of the NHLBI, and Mike Farritor for their laboratory support in obtaining the manuscript. This work was supported by the National Institutes of Health Intramural Research Program.