Supporting Information

Kapoor et al. 10.1073/pnas.1101794108

SI Materials and Methods

Genome Sequencing and Phylogenetic Analysis. Sequences with similarity to flaviviruses were assembled against prototype hepatitis C virus (HCV) strains. Gaps were filled by primer walking using specific and degenerate flavivirus primers. Both termini of the genome were acquired by using RACE (1). Thereafter, sequence validity was tested in 4× genome coverage by classical dideoxy Sanger sequencing. Nucleotide compositions of different flaviviruses and canine hepacivirus (CHV) were determined by using EMBOSS compseq (http://emboss.bioinformatics.nl/cgi-bin/emboss/compseq). Translated amino acid sequences were aligned with ClustalW. Trees were constructed by neighbor joining of pairwise amino acid distances with the program MEGA5 (2), using bootstrap resampling to determine robustness.

Screening and Quantitative PCR. All respiratory and tissue samples were extracted with Qiagen viral RNA extraction kit and RNeasy tissue DNA/RNA extraction kit. RNA was converted to cDNA using random primers and then used in nested PCR with primers for the first round (Chv-0F1: 5′-TCCACCTATGGTAAGTTC-TTAGC-3′ and Chcv-0R1: 5′-ACCCTGTCATAAGGGCGTC-3′) and the second round (Chcv-0F2: 5′-CCTATGGTAAGTTC-TTAGCTGAC-3′ and Chcv-0R2: 5′-CCTGTCATAAGGGCG-TCCGT-3′). All PCR products were sequenced to confirm the presence of CHV in samples. Quantitative PCR to determine the CHV genome copy number in respiratory samples was performed by using SYBR green chemistry and a plasmid containing HCV helicase gene as a copy number standard. The primers used were 5′-GCCATAGCACAGACTCCAC-3′ (CHV-SG-F1) and 5′-GACGGAAACATCCAAACCCCG-3′ (CHV-SG-2R1) with ready-to-use PCR mix (Applied Biosystems).

Evolutionary Analysis. Bayesian Markov chains Monte Carlo (MCMC) phylogenies and associated time to most recent common ancestor (TMRCA) for representative members of the HCV strains, CHV-01, and GHV-B were estimated by using a 555-nt segment of the NS5B gene in the program BEAST v1.6 (3). TMRCA was estimated by using a relaxed molecular clock with an uncorrelated log-normal distribution on the rate that was calibrated by using external rate estimates based on the NS5B genes of (*i*) the global diversity of HCV subtypes 1a and 1b (4) and (*ii*) HCV subtype 6 diversity in Asia (5). Normal and log-normal distributions were determined by the mean and 95% highest posterior densities (HPDs) of the reported substitution rates for all three codon positions as well as only the first and second codon positions to limit the effect of potential site saturation at the third position. A general time reversible of nucleotide substitution was used, with rate heterogeneity among sites modeled by a discrete gamma distribution with four rate categories, as determined by ModelTest (6). All analyses were performed with several tree priors, including a speciation model (Yule) and two unconstrained coalescent models, the Bayesian Skyline (7) and Bayesian Skyride (8) demographic models. MCMC sampling was performed for 5×10^7 generations, sampling every 5,000 generations. Convergence and mixing were assessed with the program Tracer v1.5 (http://tree.bio.ed.ac.uk). Maximum clade credibility trees were generated with TreeAnnotator (3).

For the data sets calibrated with both HCV subtypes 1 a/b and subtype 6, the Yule speciation model had the best fit to the data, as assessed by comparing the posterior tree likelihoods (Fig. S3). Analyses that included third-codon positions resulted in wider 95% HPDs around the mean TMRCA, likely because of an increased number of substitutions at that site. However, all model-prior combinations for each of the rate calibrations resulted in 95% HPDs that were overlapping between the analyses, indicating that estimates are robust to the choice of tree prior and inclusion of third-codon positions.

1. Kapoor A, et al. (2008) A highly prevalent and genetically diversified *Picornaviridae* genus in South Asian children. *Proc Natl Acad Sci USA* 105:20482–20487.
2. Kumar S, Nei M, Dudley J, Tamura K (2008) MEGA: A biologist-centric software for evolutionary analysis of DNA and protein sequences. *Brief Bioinform* 9:299–306.
3. Drummond AJ, Rambaut A (2007) BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol Biol* 7:214.
4. Magiorkinis G, et al. (2009) The global spread of hepatitis C virus 1a and 1b: A phylodynamic and phylogeographic analysis. *PLoS Med* 6:e1000198.
5. Pybus OG, et al. (2009) Genetic history of hepatitis C virus in East Asia. *J Virol* 83: 1071–1082.
6. Posada D, Crandall KA (1998) MODELTEST: Testing the model of DNA substitution. *Bioinformatics* 14:817–818.
7. Drummond AJ, Rambaut A, Shapiro B, Pybus OG (2005) Bayesian coalescent inference of past population dynamics from molecular sequences. *Mol Biol Evol* 22: 1185–1192.
8. Minin VN, Bloomquist EW, Suchard MA (2008) Smooth skyride through a rough skyline: Bayesian coalescent-based inference of population dynamics. *Mol Biol Evol* 25:1459–1471.
9. Krey T, et al. (2010) The disulfide bonds in glycoprotein E2 of hepatitis C virus reveal the tertiary organization of the molecule. *PLoS Pathog* 6:e1000762.
10. Whidby J, et al. (2009) Blocking hepatitis C virus infection with recombinant form of envelope protein 2 ectodomain. *J Virol* 83:11078–11089.

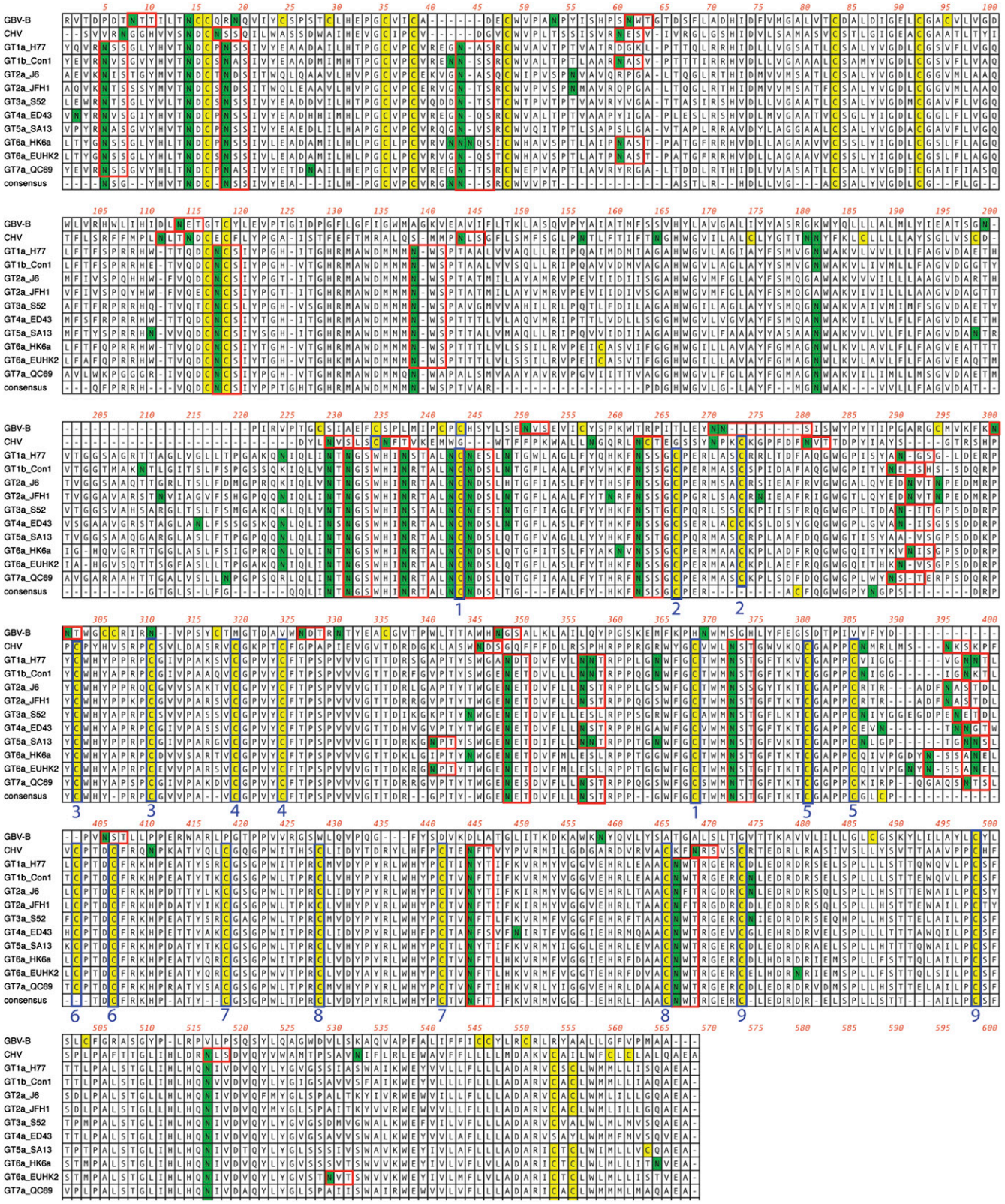

Fig. S1. Sequence alignment of envelope proteins E1 and E2 of CHV, GB virus B (GBV-B), and HCV genotypes 1a through 7a. Cysteine and asparagine residues are highlighted in yellow and green, respectively. Cysteines experimentally determined to form disulfide bridges in HCV E2 are shown in blue boxes, and blue numbers indicate disulfide connectivity (9). Predicted *N*-glycosylation sites in E1 and experimentally determined sites in E2 are shown in red boxes (10).

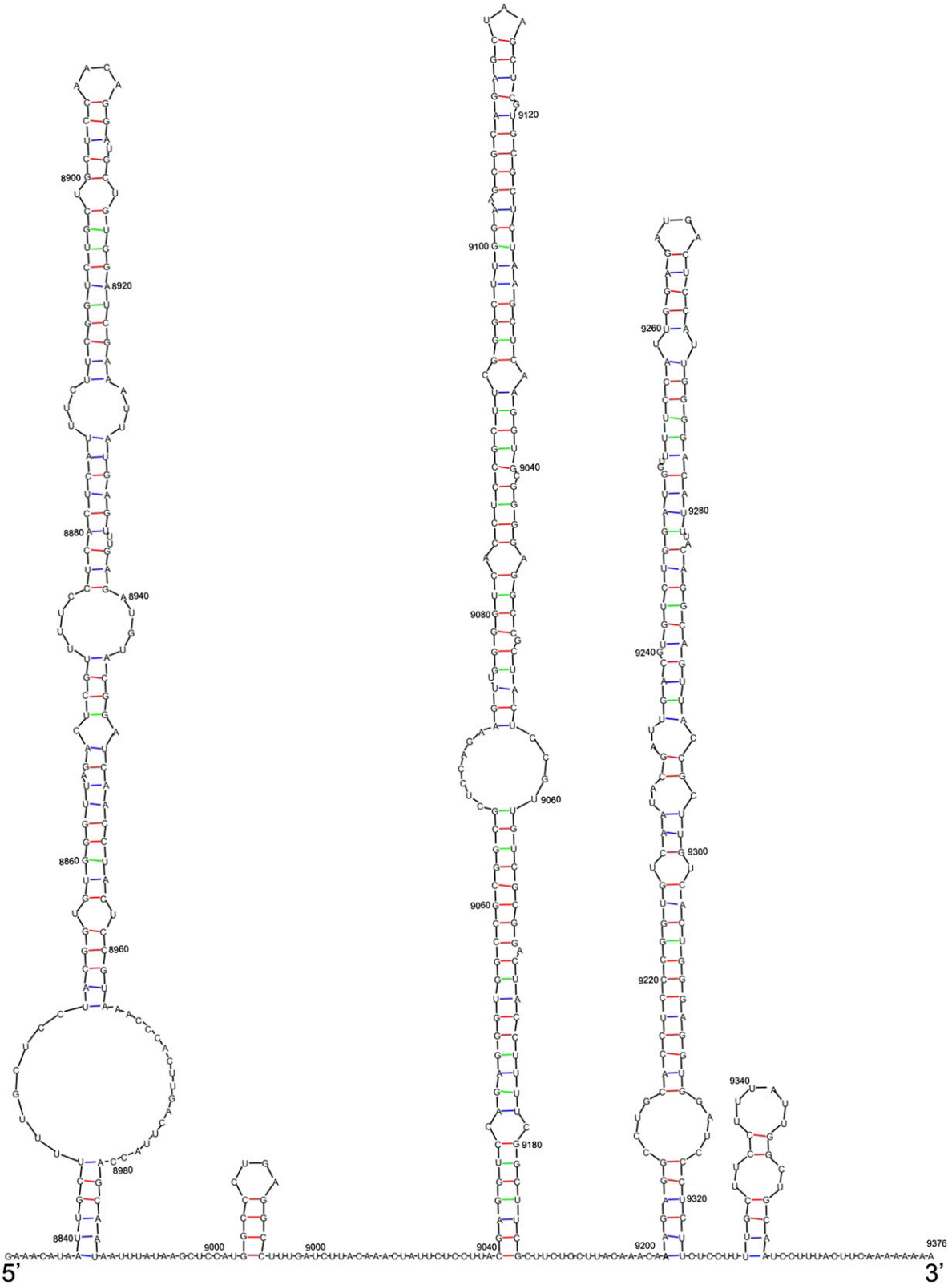

Fig. S2. RNA folding prediction with the thermodynamic folding energy minimization algorithm (MFOLD) of the terminal 540 nt of the CHV coding sequence. Base positions are numbered according to the HCV H77 numbering reference sequence.

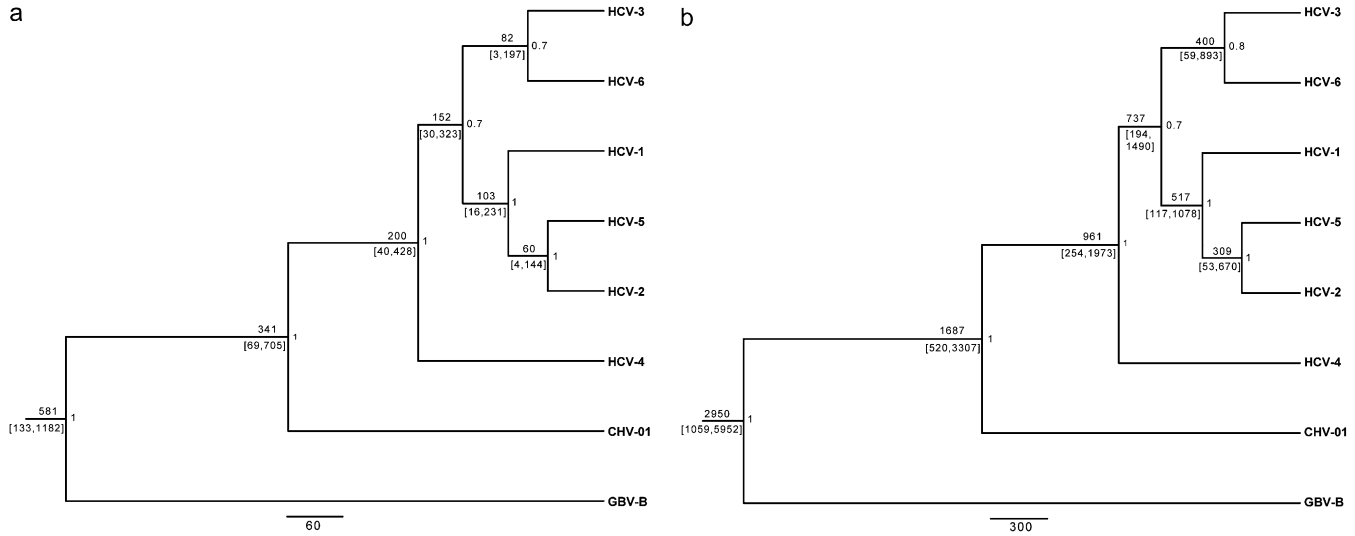

Fig. S3. Evolutionary analysis. Bayesian MCMC estimation of the TMRCA for the HCV strains, GBV-B, and CHV. Maximum clade credibility phylogeny of representative members of HCV (HCV 1: NC_004102; HCV 2: NC_009823; HCV 3: NC_009824; HCV 4: NC_009825; HCV 5: NC_009826; and HCV 6: NC_009827), hepatitis GBV-B (NC_001655), and CHV-01. TMRCAs were calculated by calibration with evolutionary rates estimated for NS5B based on HCV subtypes 1a and 1b (4) (*A*) and HCV subtype 6 (5) (*B*). The mean TMRCAs with associated 95% highest probability densities for each node are shown to the left of the node, and the Bayesian posterior probabilities are given to the right. The scale bars are in units of years before present (ybp).

Table S1. Pairwise distances between 5′ UTR, structural (S gene), and nonstructural (NS gene) proteins of different hepaciviruses

Genome region	CHCV	HCV	GBV-B	PgV
5′ UTR				
CHV	ND			
HCV	66.0	95.2		
GBV-B	56.7	62.8	ND	
PgV	IH	IH	IH	64.7
S gene				
CHV	ND			
HCV	44.1 (35.9)	67.5 (71.6)		
GBV-B	29.6 (11.2)	29.6 (12.1)	ND	
PgV	IH	IH	IH	44.8 (35.2)
NS gene				
CHV	ND			
HCV	52.3 (50.7)	66.2 (72.5)		
GBV-B	41.1 (30.2)	40.2 (30.6)	ND	
PgV	37.8 (25.5)	38.4 (25.6)	36.9 (24.3)	52.4 (49.9)

Amino acid divergence is given in parentheses. IH, Insufficient homology for valid comparison; ND, not done (only one sequence available); PgV, *Pegivirus* (GBV-A, -C, and -D).

Table S2. Sequences, accession nos., and virus abbreviations used in the phylogenetic analysis described in Fig. 4

Genus/virus	Accession no.	Description
Flavivirus		
APOIV	AF160193	Apoi virus polyprotein gene, complete cds | *Flavivirus* | Rio Bravo virus group
BANV	DQ859056	Banzi virus strain SAH 336 polyprotein gene, complete cds | *Flavivirus* | Yellow fever virus group
CHAOV	FJ883471	Chaoyang virus strain Deming polyprotein gene, complete cds | *Flavivirus*
DENV-4	AF326573	Dengue virus type 4 strain 814669, complete genome | *Flavivirus* | Dengue virus group
EHV	DQ859060	Edge Hill virus strain YMP 48 polyprotein gene | *Flavivirus* | Yellow fever virus group
GGYV	DQ235145	Gadgets Gully virus from Australia polyprotein gene, complete cds | *Flavivirus*
KADV	DQ235146	Kadam virus from Uganda polyprotein gene, complete cds | *Flavivirus*
KEDV	DQ859061	Kedougou virus strain Dak AR D1470 polyprotein gene, complete cds | *Flavivirus*
MMLV	AJ299445	Montana myotis leukoencephalitis virus complete genomic RNA | *Flavivirus*
MODV	AJ242984	Modoc virus genomic RNA for polyprotein gene | *Flavivirus* | Modoc virus group
NOUV	EU159426	Nounane virus polyprotein mRNA, complete cds | *Flavivirus*
RBV	AF144692	Rio Bravo virus strain RiMAR polyprotein gene, complete cds | *Flavivirus* | Rio Bravo virus group
SEPV	DQ859063	Sepik virus strain 7148 polyprotein gene, complete cds | *Flavivirus* | mosquito-borne viruses
SPOV	DQ859064	Spondweni virus strain SM-6 V-1 polyprotein gene | *Flavivirus* | Spondweni virus group
DENV-1	DVU88536	Dengue virus type 1 clone 45AZ5, complete genome | *Flavivirus* | Dengue virus group
KFDV	AY323490	Kyasanur forest disease virus polyprotein gene | *Flavivirus* | tick-borne encephalitis virus group
YFV	X03700	Yellow fever virus complete genome, 17D vaccine strain | *Flavivirus* | Yellow fever virus group
AEFV	AB488408	Aedes flavivirus genomic RNA, complete genome, strain Narita-21 | *Flavivirus*
CFAV	YFVCFAPP	Flavivirus cell fusing agent polyprotein gene, complete cds | *Flavivirus*
CXFV	GQ165808	Culex flavivirus strain Uganda08 polyprotein gene, partial cds | *Flavivirus*
NAKV	GQ165809	Nakiwogo virus strain Uganda08 polyprotein gene, partial cds | Viruses | Flaviviridae
Hepacivirus		
HCV-1a	AF011751	HCV strain H77 pCV-H77C polyprotein gene, complete cds
HCV-1b	HPCJCG	HCV ORF gene, complete cds | *Hepacivirus*
HCV-2b	HPCJ8G	D10988 D01221 HCV genome
HCV-2a	HPCPOLP	HCV genomic RNA for polyprotein, complete cds | *Hepacivirus*
HCV-3a	HPCEGS	HCV (isolate NZL1) genomic RNA, complete genome | *Hepacivirus*
HCV-3k	HPCJK049E1	HCV (isolate JK049) genomic RNA, complete genome | *Hepacivirus*
HCV-4a	HCV4APOLY	Y11604 HCV type 4a RNA for HCV polyprotein
HCV-5a	HCV1480	Y13184 HCV genotype 5a RNA for HCV polyprotein
HCV-6a	HCV12083	Y12083 HCV genotype 6a RNA for HCV polyprotein
HCV-6g	HPCJK046E2	HCV (isolate JK046) genomic RNA, complete genome | *Hepacivirus*
HCV-7a	EF108306	HCV (isolate QC69) polyprotein gene, complete cds | *Hepacivirus*
GBV-B	HGU22304	U22304 hepatitis GBV-B polypeptide complete genome
Pegivirus		
SPgV	AF023424	Hepatitis GB virus A complete genome
SPgV	AF023425	Hepatitis GB virus A complete genome
SPgV	HGU22303	U22303 hepatitis GB virus A polyprotein, complete cds
SPgV	HGU94421	U94421 hepatitis GB virus A strain Alab, complete genome
HPgV	AB003291	Hepatitis GB virus C genomic RNA for polyprotein, isolate CG12LC
HPgV	AB003292	Hepatitis GB virus C genomic RNA for polyprotein, isolate G05BD
HPgV	D87713	Hepatitis GB virus C genomic RNA, complete sequence, strain K2141
HPgV	HGU637155	U63715 Hepatitis GB virus C polyprotein gene, complete cds
SPgVtro	AF070476	GB virus C variant troglodytes, complete genome,
BPgV	GU566735	GB virus D strain 93 polyprotein precursor, gene, partial cds
Pestivirus		
BDV-1a	AF037405	Border disease virus strain X818, complete genome | *Pestivirus*
BVDV-1a	BVDCG	Bovine viral diarrhea virus 1-NADL, complete genome | *Pestivirus*
BVDV-2	AF002227	Border disease virus strain C413, complete genome | *Pestivirus*
CSFV-1	HCVCG3PE	Classical swine fever virus, Brescia hog cholera virus protein precursor | *Pestivirus*
Gir-PV	AF144617	Pestivirus giraffe-1 H138 complete genome | *Pestivirus*
BDV-4	GU270877	Border disease virus strain H2121 (Chamois-1), complete genome | *Pestivirus*
	FJ040215	Bovine viral diarrhea virus 3 Th/04 KhonKaen, complete genome | *Pestivirus*
Unassgd	EF100713	Porcine pestivirus isolate Bungowannah polyprotein gene, partial cds | *Pestivirus*

cds, coding sequence; Unassgd, unassigned.