Supporting Information

Kapoor *et al.* 10.1073/pnas.0807979105

SI Text

Determination of 5′ UTR Structure. In the deduced type II IRES stem loops H, I, K, and L (as classified in ref. 1) and unpaired regions showed nucleotide sequence similarity or identity to sequences from the other four most closely related picornavirus genera (Fig. S3*B* gray boxes, labeled 1–14). HCoSV-A1 additionally contained the short polypyrimidine (Py) tract (Fig. S3 *A* and *B*, gray box labeled 15) immediately downstream of the last stem loop, L, and showed a similar spacing upstream from the predicted start codon of the HCoSV-A1 long ORF. Although there was little or no sequence identity in the duplex stems on these structures between virus groups (Fig. S3*B*), the H, I, K, and L stem loops were of very similar size and relative spacing.

The main structural difference between HCoSV-A1 and a typical type II IRES was the existence of an extremely long, bifurcated stem loop between the I and K structures (Fig. S3*A*). In other type II IRESs, this region, designated stem-loop J (or domain 4–1) contains a single stem loop with lengths ranging from 37–50 bases in FMDV and ERAV, 55 bases in erboviruses, and 37–43 in cardioviruses. The length of the corresponding region in HCoSV-A1 was 140 bases. No detectable sequence homology in the terminal loop or duplex stem sequences of HCoSV-A1 was found with equivalent regions in other picornaviruses. However, HCoSV-A1 retains a short region of nucleotide sequence homology in the duplex region in the proximal stem (labeled 10 and 11) and the A-bulge (conserved region 13; Fig. S3*B*), shown to be the specific regions that bind to the ribosomal eIF4B component (2, 3).

Although structures I–L represent the main conserved elements in type II IRESs, HCoSV-A1 showed evidence for additional RNA secondary structures upstream of the predicted IRES, including putative homologues of stem loops F and G, identified in cardioviruses (4) and a possible homologue of stem loop D upstream from a short, second interrupted predicted unpaired polypyrimidine tract (positions 551–575). These HCoSV-A1 structures showed similar spacings upstream from structure I in cardioviruses and contained identical (D) or similar (F) terminal unpaired regions (tetraloops and pentaloops respectively). Sequence variation in both strands of the F stem maintained absolute complementarity supporting the proposed F stem loop structure. The small stem loop upstream of H showed an identical pentaloop sequence to a structure in the equivalent position in cardioviruses (position 398). Where determined, the region encompassing stem loops D–G does not contribute to IRES function and is known to be highly variable between different genera. No comparable structures are identifiable in erbovirus or ERAV sequences (5, 6), whereas in FMDV, the region contains a large stem loop that forms the *cis*-replicating element (7).

The 3′end of the HCoSV-A1 genome was polyadenylated (A_{29}). Excluding this, the 3′UTR was 95 bases in length and contained a consecutive series of 13 U_{2-4} sequences interspersed with almost invariably purines (R_{1-3}). Although this region is known to structurally variable between picornavirus genera and precludes the type of structure modeling carried out on the 5′UTR, MFOLD did identify a thermodynamically stable large terminal stem loop from position 7522 to 7611, containing a single long interrupted duplex that incorporated the stop codon at position 7537 (data not shown).

Finally, the HCoSV-A1 sequence was analyzed for evidence of genome-scale ordered RNA structure (GORS) by comparison of folding energies of consecutive fragments of nucleotide sequence with sequence order randomized controls (8). In contrast to its detection in aphthoviruses, erboviruses, cardiovirus, and teschoviruses (mean MFEDs for coding regions 4.1–8.5%), there was no sequence order-dependent structure in the coding region of the HCoSV-A1 genome (mean MFED −1.2%), a finding shared with the sole member of other genus related to HCoSV-A1, SVV (mean MFED 0.1%). Absence of GORS suggests that HCoSV-A1 infections may be acute and resolving, reducing the likelihood that chronic infections with HCoSV-A1 might occur (8).

Prevalence of HEV Infections. We tested 41 AFP cases and all 41 healthy controls for the presence of HEV by using RT-nested PCR with primers targeting the 5′ UTR. All positive PCRs were confirmed as HEV-B or HEV-non B by sequencing; 31/41 AFP and 25/41 control stools were HEV positive, showing an equal rate of HEV detection in both groups ($P = 0.15$). Coinfections with HEV and HCoSV were no more frequent in AFP cases than controls. Significantly more HEV than HCoSV infections were detected in AFP (76% v 49%; $P = 0.008$) but not in healthy children (61% v 44%; $P = 0.12$). HEV-positive children were younger than HEV-negative children (45.3 months versus 76.3 months, $P = 0.04$) and more likely to have fever at onset of AFP (25/31 with fever versus 6/31 without, $P = 0.023$) and had the same male/female ratio. Whether HEV infected or not AFP cases showed the same clinical profile relative to paralytic asymmetry, progression at 2 weeks follow-up or total number of oral poliovirus vaccine doses previously received. We determined whether HCoSV and HEV coinfections were more frequent than expected assuming independent events. Infection with 1 virus genus was significantly associated with infection by the other virus ($P = 0.01$) possibly reflecting related transmission routes.

Picornavirus Sequences for Phylogenetic Analysis. For determination of sequence divergences between cosaviruses and those of other picornaviruses, sequences were aligned from 2 or more representative serotypes from each species within each of the 9 currently classified genera of picornaviruses, members of the proposed "*Sapelovirus*" genus, and the currently unclassified viruses duck hepatitis virus (DHV) (9), Seneca Valley virus (SVV), and seal picornavirus (SePV-1) (10). Sequences used for the comparison comprised the following: Genus *Aphthovirus*, foot-and-mouth disease (FMDV) NC_011450, FMDVALF, FDI320488, NC_004004, NC_002554, FMV7572, NC_004915, AY687334, NC_003992, FDI251473, AY593843, NC_011451, AY593853, NC_011452, equine rhinitis A virus (ERAV) NC_003982, ERVPOLY; genus *Enterovirus* Simian enterovirus NC_003988, human enterovirus, species C NC_002058, HPO132960, POL2LAN, POL544513, POL3L37, HPO293918, NC_001428, CXA24CG, species B NC_002347, NC_001360, NC_000881, NC_001657, NC_002601, NC_001342, NC_001656, NC_001472, species A NC_001612, ETU22522, species D NC_001430, AY426531, bovine enterovirus BEVVG527, AY508697, human rhinoviruses NC_001617, DQ473500, EF173418, DQ473490, EF173420, EF186077, EF077279, EF077280, NC_001490; genus *Hepatovirus*: avian encephalomyocarditis virus (AEV) NC_003990, AJ225173, AY517471, AY275539, simian hepatitis A virus (HAV) SHVAGM27, human HAV HAVRNAGBM, NC_001489; genus *Cardiovirus*: encephalomyocarditis virus (EMCV) NC_001479, XXEVCG, Theiler's virus NC_001366, AB090161, Saffold virus EF165067,

AM922293; genus *Erbovirus*: NC_003077, NC_003983; genus *Parechovirus*: Ljungan virus AF538689, AF327921, AF327922, human parechoviruses (HPeV) ECHPICORN, NC_001897, AB084913, AF055846, NC_008286, AB252582; genus *Kobuvirus*: bovine kobuvirus NC_004421 and Aichi virus NC_001918 and DQ028632; Genus *Teschovirus*: PEN011380, AF296088, AF296107, AF296112, AF296090, AF296092, AF296093, AF296094, AF296119, AF296117; currently unclassified viruses DHV: EF382778, DQ249301, DQ249300, DQ249299, EF093502, SVV: DQ641257, SePV-1: NC_009891 and members of the provisionally assigned genus "*Sapelovirus*": porcine enterovirus A NC_003987, simian picornavirus 1 AY064708 and duck picornavirus NC_006553.

For comparison with human enterovirus diversity, the following datasets were used: HEV species B: AY302539, NC_000873, DQ246620, DQ534205, AY302558, AF465517, NC_001360, AF311939, EF174468, EF174469, AF085363, NC_000881, CXU57056, AY752945, AF231763, AY673831, AY875692, AF114383, NC_001342, AY167104, AY167105, E11577590, E11577589, E11577594, AY167106, E11276224, EV11VPCD, AY036578, AY302559, AF524867, AF524866 and NC_001656; for interspecies comparisons and intergenus comparisons (with the sapelovirus sequences listed above), the following sequences of HEV species A-D were used: DQ443002; AB205396; POL3L37; POL2LAN; V01148; AY790926; AB204853; AB192877; AY421763; AY421764; NC_001430; AY426531, NC_001656, NC_002347, EV11VPCD, DQ534205 and NC_001342. All alignments were numbered by using POL3L37 as a reference sequence.

Picornavirus Sequences for Genome-Scale Ordered RNA Structure Comparisons. The following complete genome sequences were used: for the genus Aphthovirus, GenBank accession nos. AF154271, FMDVALF, NC_002554, NC_003982, NC_003992, NC_004004, and NC_002527 for FMDV; for the genus Cardiovirus, accession nos. NC_001479 and MNG POLY for encephalomyocarditis virus and NC_001366 for Theiler's virus; for the genus Enterovirus, accession nos. NC_001428, NC_001430, NC_001472, NC_001490, NC_001612, NC_001617, NC_001752, NC_001859, NC_002058, NC_003986, NC_003988, POL3L37, and SVDMPS; for the genus Erbovirus, accession no. NC_003983; for the genus Hepatovirus, accession nos. NC_003990 for avian encephalomyocarditis virus, SHVAGM27 for simian HAV, and NC_001489 for human HAV; for the genus Kobuvirus, accession nos.s NC_004421 for bovine kobuvirus and NC_001918 for Aichi virus; and for the genus Teschovirus, accession nos. AB038528, AF231769, AF296087, AF296091, AF296093, AF296115, AF296119, and NC_003985.

1. Duke GM, Hoffman MA, Palmenberg AC (1992) Sequence and structural elements that contribute to efficient encephalomyocarditis virus RNA translation. *J Virol* 6:1602–1609.
2. Lopez de Quinto S, Martinez-Salas E (2000) Interaction of the eIF4G initiation factor with the aphthovirus IRES is essential for internal translation initiation in vivo. *RNA* 6:1380–1392.
3. Saleh L, *et al.* (2001) Functional interaction of translation initiation factor eIF4G with the foot-and-mouth disease virus internal ribosome entry site. *J Gen virol* 82:757–763.
4. Hellen CU, de Breyne S (2007) A distinct group of hepacivirus/pestivirus-like internal ribosomal entry sites in members of diverse picornavirus genera: Evidence for modular exchange of functional noncoding RNA elements by recombination. *J Virol* 81:5850–5863.
5. Hinton TM, Crabb BS (2001) The novel picornavirus equine rhinitis B virus contains a strong type II internal ribosomal entry site which functions similarly to that of Encephalomyocarditis virus. *J Gen Virol* 82:2257–2269.
6. Hinton TM, Li F, Crabb BS (2000) Internal ribosomal entry site-mediated translation initiation in equine rhinitis A virus: Similarities to and differences from that of foot-and-mouth disease virus. *J Virol* 74:11708–11716.
7. Mason PW, Bezborodova SV, Henry TM (2002) Identification and characterization of a cis-acting replication element (cre) adjacent to the internal ribosome entry site of foot-and-mouth disease virus. *J Virol* 76:9686–9694.
8. Simmonds P, Tuplin A, Evans DJ (2004) Detection of genome-scale ordered RNA structure (GORS) in genomes of positive-stranded RNA viruses: Implications for virus evolution and host persistence. *RNA* 10:1337–1351.
9. Tseng CH, Knowles NJ, Tsai HJ (2007) Molecular analysis of duck hepatitis virus type 1 indicates that it should be assigned to a new genus. *Virus Res* 123:190–203.
10. Kapoor *et al.* (2008) A highly divergent picornavirus in a marine mammal. *J Virol* 82:311–320.

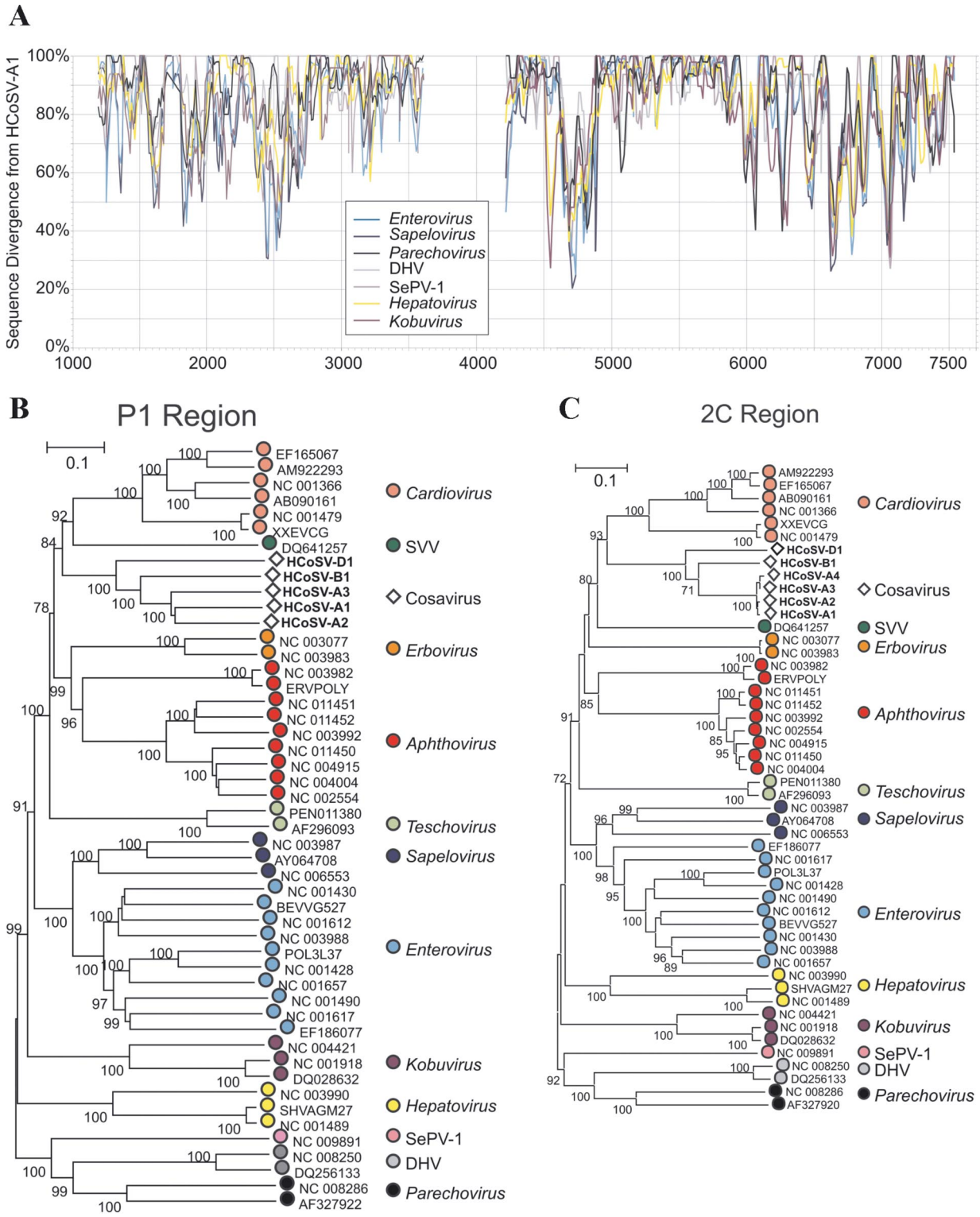
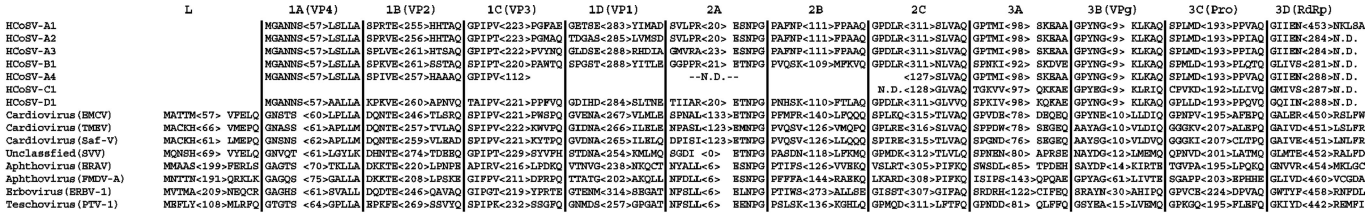


Fig. S1. Sequence similarity detection and phylogenetic relationships of cosaviruses with other picornaviridae genera. (*A*) Mean divergence calculated for a sliding window of pairwise translated protein *p* distances between HCoSV-A1 and the genera of the *Picornaviridae* family not included in Fig. 1*A*. (*B* and *C*) Phylogenetic analysis of the P1 (*B*) and 2C (*C*) regions with representative of each picornavirus genera (two sequences per species or genus) by neighbor-joining using amino acid *p* distances.



	L	1A(VP4)	1B(VP2)	1C(VP3)	1D(VP1)	2A	2B	2C	3A	3B(VPg)	3C(Pro)	3D(RdRp)
HCoSV-A1		MGANNS<57>LSLLA	SPRTE<255>HHTAQ	GPIPV<223>PGFAE	GETSE<283>YIMAD	SVLPR<20>	ESNPG	PAPNP<111>FPAAQ	GPDLR<311>SLVAQ	GPTMI<98>	SKEAA	GPYNG<9> KLKAQ	SPLMD<193>PPVAQ	GIIEN<453>NKLSA
HCoSV-A2		MGANNS<57>LSLLA	SPRVE<256>HHTAQ	GPIPT<223>PGMAQ	TDGAS<285>LVMSD	SVLPR<20>	ESNPG	PAPNP<111>FPAAQ	GPDLR<311>SLVAQ	GPTMI<98>	SKEAA	GPYNG<9> KLKAQ	SPLMD<193>PPIAQ	GIIEN<284>N.D.
HCoSV-A3		MGANNS<57>LSLLA	SPLVE<261>HTSAQ	GPIPT<222>PVYNQ	GLDSE<288>RHDIA	GMVRA<23>	ESNPG	PAPNP<111>FPAAQ	GPDLR<311>SLVAQ	GPTMI<98>	SKEAA	GPYNG<9> KLKAQ	SPLMD<193>PPIAQ	GIIEN<284>N.D.
HCoSV-B1		MGANNS<57>LSLLA	SPKVE<261>HSTAQ	SPIPT<220>PAWTQ	SPGST<288>YITLE	GGPPR<21>	ETNPG	PVQSK<109>MPKVQ	GPDLR<311>NLVAQ	SPNKI<92>	SKDVE	GPYNG<9> KLKAQ	SPMLD<193>PLQTQ	GLIVR<281>N.D.
HCoSV-A4		MGANNS<57>LSLLA	SPIVE<257>HAAAQ	GPIPV<112>		--N.D.--			<127>SLVAQ	GPTMI<98>	SKEAA	GPYNG<9> KLKAQ	SPLMD<193>PPVAQ	GIIEN<288>N.D.
HCoSV-C1									N.D.<128>GLVAQ	TGKVV<97>	QKEAE	GPYEG<9> KLRIQ	CPVKD<192>LLIVQ	GMIVS<287>N.D.
HCoSV-D1		MGANNS<57>AALLA	KPKVE<260>APNVQ	TAIPV<221>PPFVQ	GDIHD<284>SLTNE	TIIAR<20>	ETNPG	PNHSK<110>FTLAQ	GPDLR<311>GLVVQ	SPKIV<98>	KQKAE	GPYNG<9> KLKAQ	GPLLD<193>PPQVQ	GQIIN<288>N.D.
Cardiovirus(EMCV)	MATTM<57> VFELQ	GNSTS <60>LPLLA	DQNTE<246>TLSRQ	SPIPV<221>PWSPQ	GVENA<267>VLMLE	SPNAL<133>ETNPG	PFMFR<140>LPQQQ	SPLKQ<315>TLVAQ	GPVDR<78>	DEQEQ	GPYNE<10>LLDIQ	GPNPV<195>AFEPQ	GALER<450>RSLFW
Cardiovirus(TMEV)	MACKH<66> VMEPQ	GNASS <61>APLLM	DQNTE<257>TVLAQ	SPIPV<222>KWVPQ	GIDNA<266>ILELE	NPASL<123>EMNPG	PVQSV<126>VMQPQ	GPLRE<316>SLVAQ	SPPDW<78>	SEGEQ	AAYAG<10>VLDIQ	GGGKV<207>ALEPQ	GAIVD<451>LSLFR
Cardiovirus(Saf-V)	MACKH<61> LMEPQ	GNSNS <62>APLLM	DQNTE<259>VLEAD	SPIPV<221>KYTPQ	GVDNA<265>ILKLQ	DPISI<123>ETNPG	PVQSV<126>LLQQQ	SPIRE<315>TLVAQ	SPGND<76>	SEGEQ	AAYSG<10>VLDVQ	GGGKI<207>CLTPQ	GAIVE<451>LNLFR
Unclassfied(SVV)	MQNSH<69> VYELQ	GNVQT <61>LGYLK	DHNTE<274>TDEBQ	GPIPT<229>SYVPH	STDNA<254>KMLMQ	SGDI <0>	ETNPG	PASDN<118>LFKMQ	GPMDK<312>TLVLQ	SPNEN<80>	APRSE	NAYDG<12>LMEMQ	QPNVD<201>LATMQ	GLMTE<452>RALFD
Aphthovirus(ERAV)	MMAAS<199>FERLS	GAGTS <70>TKLLA	DKETE<220>LPNPE	APIRV<216>LPDKQ	VTNVG<238>NKQCT	NYALL<6>	ESNPG	PTIFS<126>VVSKQ	VSLRT<305>PIFKQ	SWSDL<85>	TPDEH	SAYDP<14>KIRTE	TGVPA<195>LPQKQ	GNVVR<454>MKLGC
Aphthovirus(FMDV-A)	MNTTN<191>QRKLK	GAGQS <75>GALLA	DKETE<208>LPSKE	GIFPV<211>DPRPQ	TTATG<202>AKQLL	NFDLL<6>	ESNPG	PFFFA<144>RAEKQ	LKARD<308>PIFKQ	ISIPS<143>QPQAE	GPYAG<61>LIVTE	SGAPP<203>EPHHE	GLIVD<460>VCGDA
Erbovirus(ERBV-1)	MVTMA<209>NEQCR	GAGHS <61>SVALL	DQDTE<246>QAVAQ	GIPGT<219>YPRTE	GTENM<314>SEGAT	NFSLL<6>	ELNPG	PTIWS<273>ALLSE	GISST<307>GIFAQ	SRDRH<122>CIPEQ	SRAYN<30>AHIPQ	GPVCE<224>DPVAQ	GWTYF<458>RNPDL
Teschovirus(PTV-1)	MSFLY<108>MLRPQ	GTGTS <64>GPLLA	EPKFE<269>SSVTQ	SPIPK<232>SSGPQ	GNMDS<257>GPGAT	NFSLL<6>	ESNPG	PSLSK<136>KGHLQ	GPMQD<311>LFTPQ	GPNDD<81>	QLFPQ	GSYEA<15>LVEMQ	GPKGQ<195>PLEPQ	GKIYD<442>REMFI

Fig. S2. Predicted location of polyprotein cleavage sites. The cleavage sites are indicated by vertical bars. The length of the intervening protein is shown in brackets. Genomes used were cardiovirus Saf-V(EF165067), EMCV (M81861), TMEV (M20562), ERAV (NC_003982), FMDV-A (NC_011450), SVV (DQ641257), ERBV-1 (NC_003983), and PTV-1 (NC_003985). N.D. indicates regions not determined.

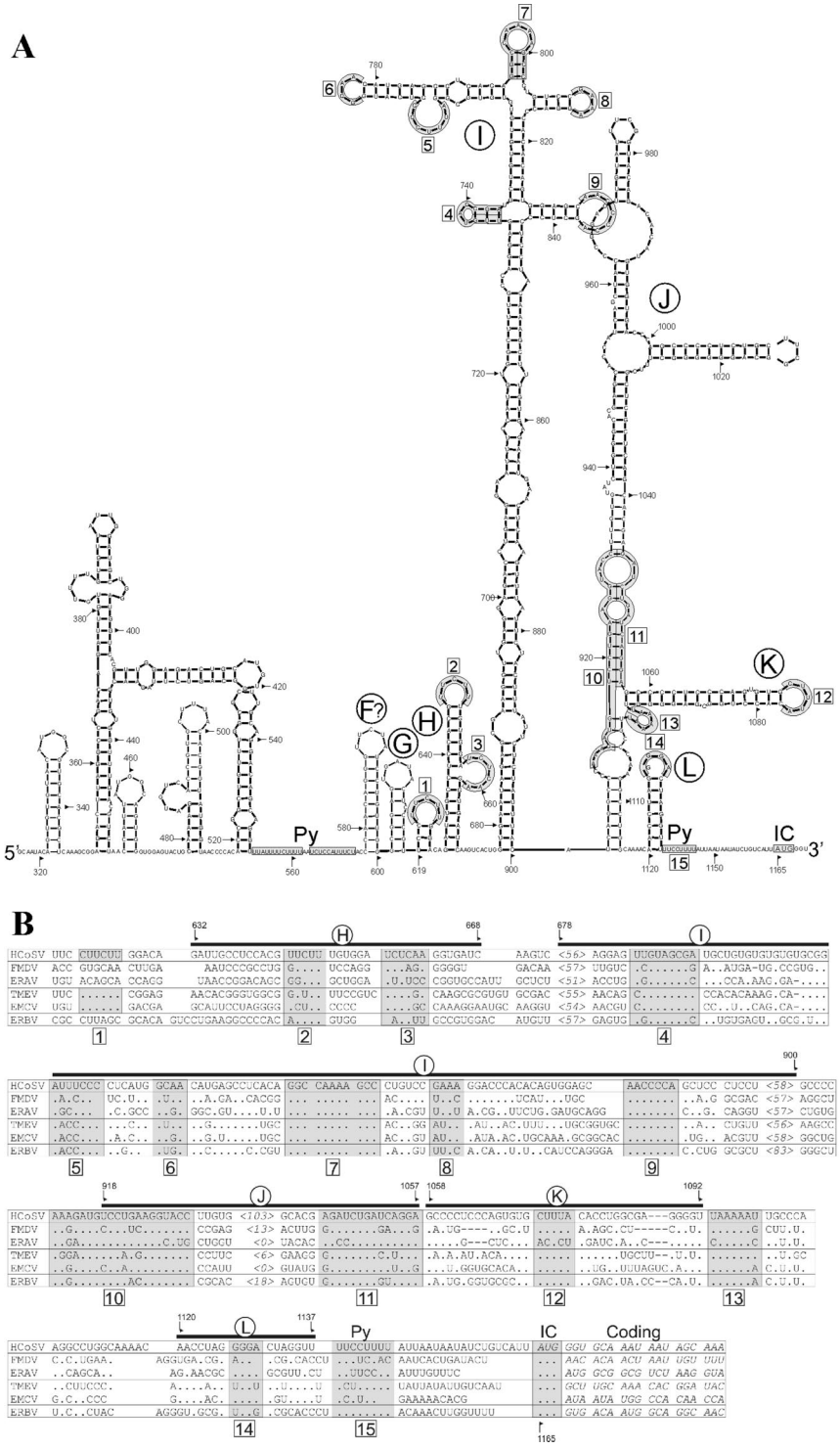


Fig. S3. Proposed secondary structure of the HCoSV-A1 5′UTR RNA and conserved regions in type II IRES. (*A*) Folding was predicted by MFOLD and by modeling onto currently described IRES structures of cardioviruses, FMDV, ERAV and erboviruses. RNA structures in the IRES have been labeled as described (1). Stem loops H, I, J/K, and L correspond to domains 2–5 of the FMDV IRES (7). Regions of sequence similarity or identity with other picornavirus type II IRES sequences are indicated by shaded boxes (numbered 1–15). IC, initiating codon; Py, polypyrimidine tract. (*B*) Alignment of the predicted IRES region of HCoSV-A1 with other type II IRES containing picornaviruses (FMDV: NC_011450; ERAV: NC_003982; TMEV: NC_001366; EMCV: NC_001479; ERBV: NC_003077). Regions of identifiable homology have been shaded and numbered 1–15 as in *A*. The spans of the stem loop structures H–L are shown as solid bars above alignments. For presentation purposes, long stretches with little or no detectable sequence homology between virus sequences have been excluded, and their length indicated in parentheses.

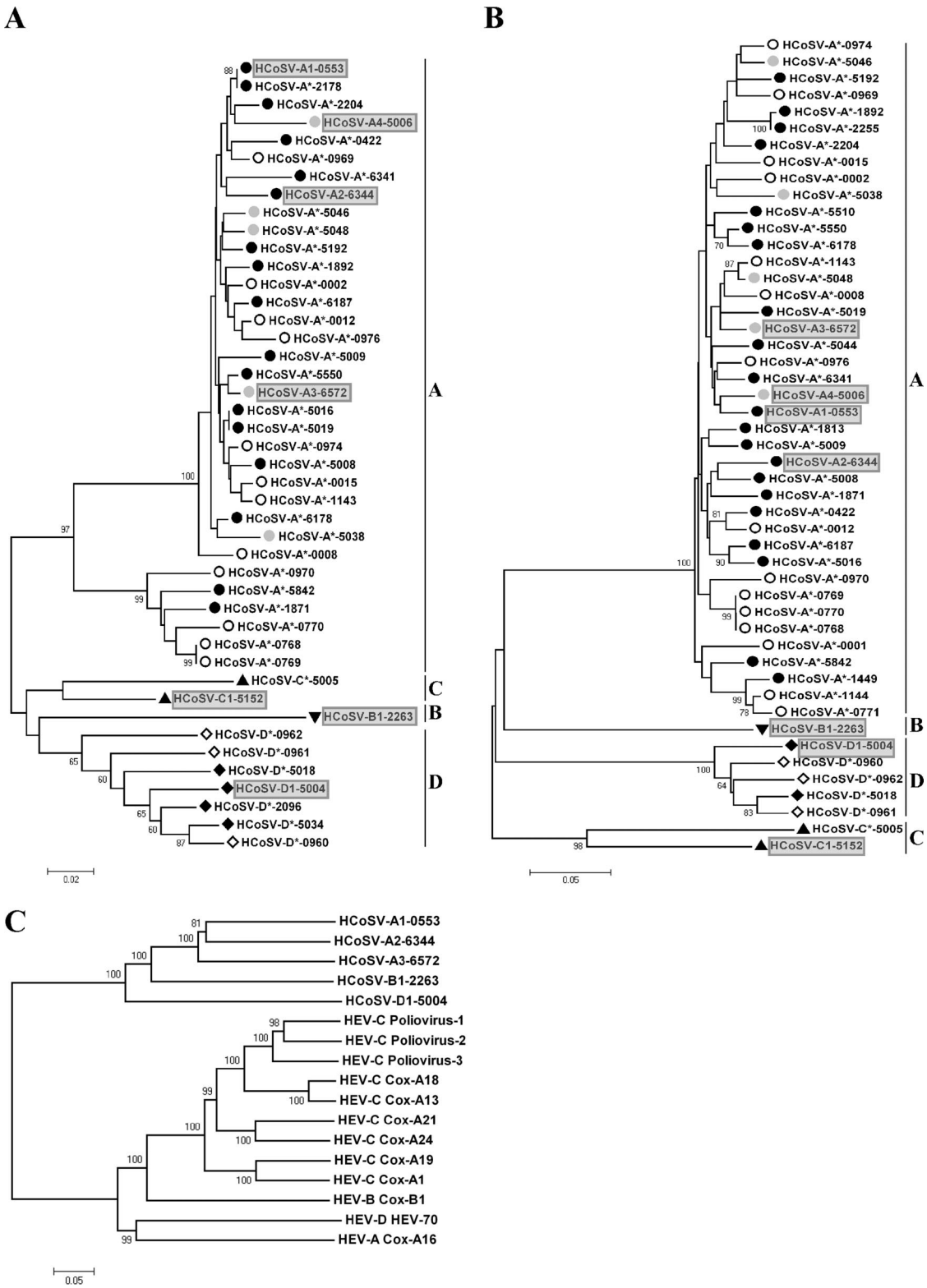


Fig. S4. Phylogenetic analyses of cosavirus 5′ UTR and 3Dpol diagnostic PCR regions, and translated P1 regions. Filled symbols are from AFP cases, empty symbols from healthy controls, and gray symbols from healthy AFP contacts. (*A*) Analysis of 5′ UTR of cosaviruses. (*B*) Analysis of 3D pol region of cosaviruses. Cosavirus species A–D shown by vertical lines. Different species of HCoSV are shown with differently shaped symbols. In gray boxes are viruses whose longer genome sequences are analyzed in Fig. 1*B* and Fig. S1 *B* and *C*. (*C*) Analysis of P1 capsid region with representative of HEV species A–D and HEV-B serotypes together with cosaviruses to highlight similar branch lengths in the HEV and Cosavirus genera.

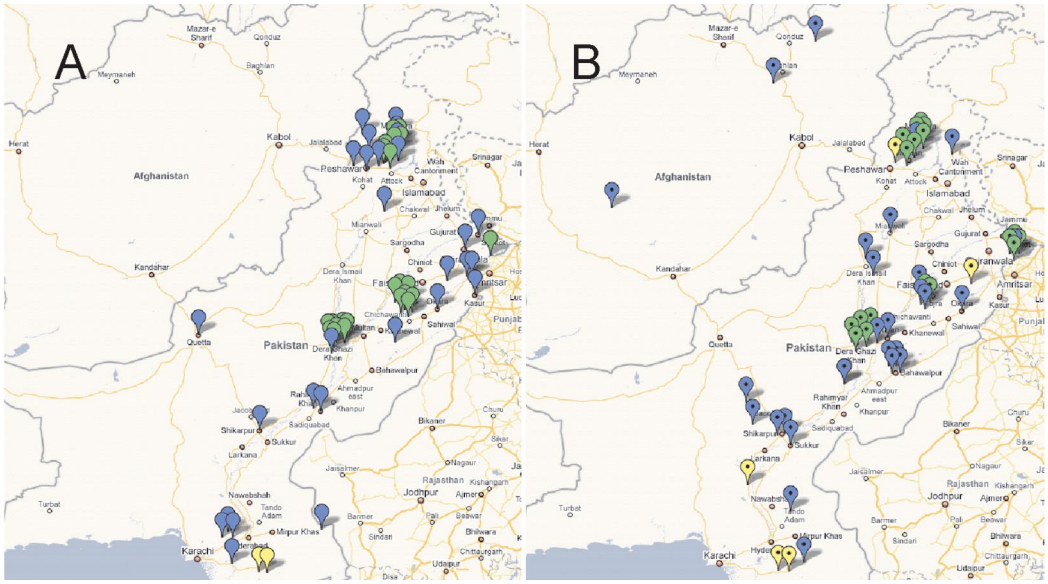


Fig. S5. Geographic distribution of stool samples analyzed. Collection site of cosavirus negative (*A*) and positive (*B*) samples. Green, blue, and yellow markers are for healthy, AFP, and contact children, respectively.

Table S1. Cosavirus sequence divergence

P1 Region

Nucleotide similarity

Group	A	B	D	
A	67.0%	64.7%	57.4%	53.5%
B	55.3%	- - -	- - -	56.0%
D	47.6%	49.4%	- - -	- - -
	A	B	D	

Amino acid similarity

3D Region

Nucleotide similarity

Group	A	B	D	C	
A	96.9%	93.4%	65.1%	65.8%	62.2%
B	71.6%	- - -	- - -	65.0%	65.9%
D	66.8%	68.9%	- - -	- - -	65.3%
C	62.8%	67.4%	65.6%	- - -	- - -
	A	B	D	C	

Amino acid similarity