Contents lists available at ScienceDirect

Virology

journal homepage: www.elsevier.com/locate/yviro

Group-specific structural features of the 5′-proximal sequences of coronavirus genomic RNAs

Shih-Cheng Chen, René C.L. Olsthoorn [*]

Leiden Institute of Chemistry, Department of Molecular Genetics, Leiden University, PO Box 9502, 2300 RA Leiden, The Netherlands

ARTICLE INFO

Article history:
Received 22 December 2009
Returned to author for revision
6 January 2010
Accepted 5 February 2010
Available online 4 March 2010

Keywords:
Coronavirus
SARS-CoV
RNA secondary structure
Packaging signal
5′ Untranslated region

ABSTRACT

Global predictions of the secondary structure of coronavirus (CoV) 5′ untranslated regions and adjacent coding sequences revealed the presence of conserved structural elements. Stem loops (SL) 1, 2, 4, and 5 were predicted in all CoVs, while the core leader transcription-regulating sequence (L-TRS) forms SL3 in only some CoVs. SL5 in group I and II CoVs, with the exception of group IIa CoVs, is characterized by the presence of a large sequence insertion capable of forming hairpins with the conserved 5′-UUYCGU-3′ loop sequence. Structure probing confirmed the existence of these hairpins in the group I *Human coronavirus-229E* and the group II *Severe acute respiratory syndrome coronavirus* (SARS-CoV). In general, the pattern of the 5′ *cis*-acting elements is highly related to the lineage of CoVs, including features of the conserved hairpins in SL5. The function of these conserved hairpins as a putative packaging signal is discussed.

Introduction

The emergence of the *Severe acute respiratory syndrome coronavirus* (SARS-CoV) in 2003 has boosted related research and led to the discovery of many novel coronaviruses (CoVs) from different hosts such as equines, whales, birds, and bats; the latter species are considered as the potential reservoir of SARS-CoV (Guan et al., 2003, Ksiazek et al., 2003; Li et al., 2005; Marra et al., 2003; Mihindukulasuriya et al., 2008; Woo et al., 2007, 2009; Zhang et al., 2007). In the past few years, also two novel human CoVs, NL63 and HKU1, have been identified causing rather severe symptoms in infants and the elderly (van der Hoek et al., 2004; Woo et al., 2005). The discovery of so many novel CoVs calls for a better understanding of the phylogeny of CoVs.

Based on serological patterns and genome organization, the genus *Coronavirus* has been classified into three major groups: group I, II and III (Lai and Cavanagh, 1997; Brian and Baric, 2005). More recently, these groups have been further subdivided into, in total, 9 subgroups, based upon amino acid similarity of structural and non-structural proteins (nsp) (Snijder et al., 2003; Woo et al., 2006, 2007; Woo et al., 2006, 2007). However, other studies propose at least 5 distinct lineages (Tang et al., 2006; Dong et al., 2007; Vijaykrishna et al., 2007), and even for SARS-CoV there is discussion whether it represents a separate lineage (Rota et al., 2003) or is an early split-off of group II CoVs (Snijder et al., 2003; Gibbs et al., 2004). Thus, in addition to the conventional pair-wise comparison of viral protein sequences, other genetic or structural features may be helpful in the classification of CoVs.

In the genome of CoVs, like that of most RNA viruses, the 5′ and 3′ untranslated regions (UTRs) usually harbor important structural elements which are involved in replication and/or translation (Chang et al., 1994; Raman et al., 2003; Raman and Brian, 2005; Goebel et al., 2007; Züst et al., 2008; Liu et al., 2009). In *Mouse hepatitis virus* (MHV), a group II CoV, a bulged stem–loop and a pseudoknot structure were identified in the 3′ UTR (Goebel et al., 2004a). Similar pseudoknot structures were found in other group I and II CoVs, showing structural conservations of the CoV 3′ UTR (Goebel et al., 2004a). However, the 3′ UTR of MHV could be functionally replaced by the 3′ UTR of group II SARS-CoV but not by that of the group I *Transmissible gastroenteritis virus* (TGEV) or the group III *Avian infectious bronchitis virus* (IBV), indicating certain group-specific functions for the 3′ UTR (Goebel et al., 2004b).

In this study the secondary structures of the 5′ UTRs and the 5′-proximal sequences of the ORF1ab gene in all known CoVs were predicted. The structural features of this region turned out to reflect the known grouping of CoVs, which is based on amino acid similarity. The unique and conserved features were further investigated in detail.

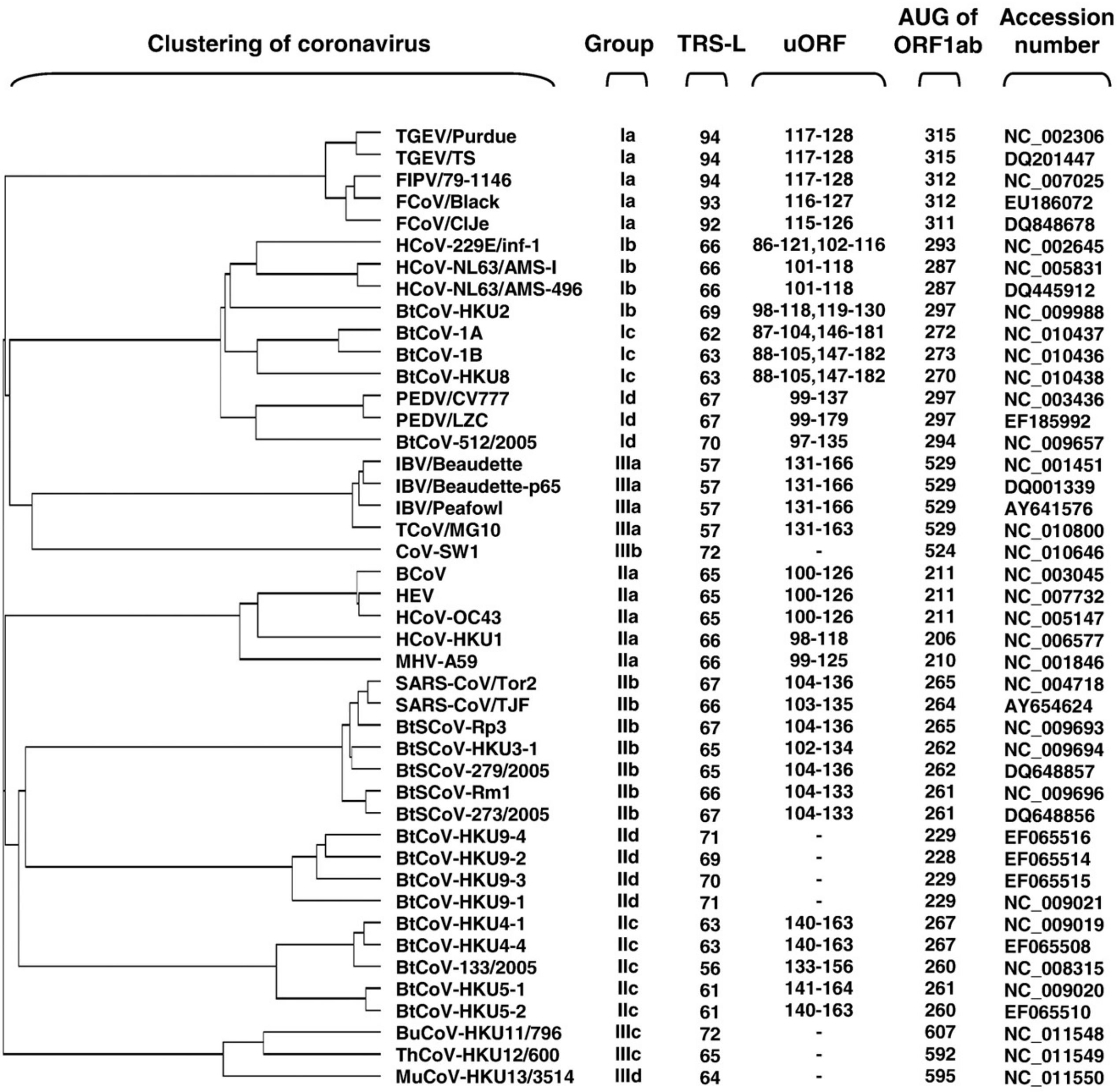
Results and discussion

The clustering of the 5′-proximal sequence of CoV RNAs shows group specificity

The clustering of the CoV 5′-proximal 420 nucleotides (nts) obtained from the *Kalign* webserver (see Materials and methods)

[*] Corresponding author. Leiden Institute of Chemistry, Department of Molecular Genetics, Gorlaeus Laboratories, Einsteinweg 55, 2333 CC, Leiden, The Netherlands. Fax: +31 715274357.
E-mail address: olsthoor@chem.leidenuniv.nl (R.C.L. Olsthoorn).

0042-6822/$ – see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.virol.2010.02.007

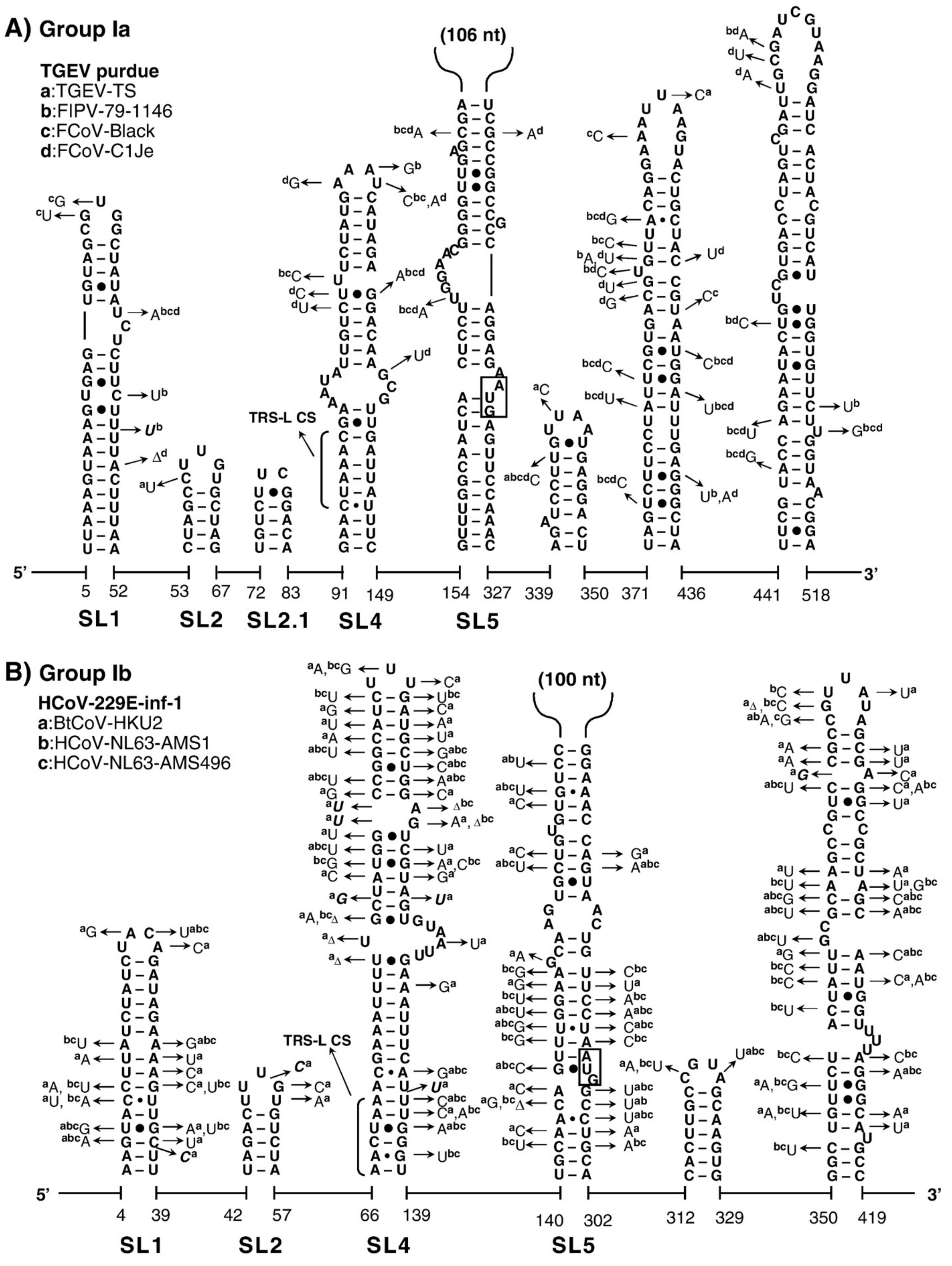


Clustering of coronavirus	Group	TRS-L	uORF	AUG of ORF1ab	Accession number
TGEV/Purdue	Ia	94	117-128	315	NC_002306
TGEV/TS	Ia	94	117-128	315	DQ201447
FIPV/79-1146	Ia	94	117-128	312	NC_007025
FCoV/Black	Ia	93	116-127	312	EU186072
FCoV/ClJe	Ia	92	115-126	311	DQ848678
HCoV-229E/inf-1	Ib	66	86-121,102-116	293	NC_002645
HCoV-NL63/AMS-I	Ib	66	101-118	287	NC_005831
HCoV-NL63/AMS-496	Ib	66	101-118	287	DQ445912
BtCoV-HKU2	Ib	69	98-118,119-130	297	NC_009988
BtCoV-1A	Ic	62	87-104,146-181	272	NC_010437
BtCoV-1B	Ic	63	88-105,147-182	273	NC_010436
BtCoV-HKU8	Ic	63	88-105,147-182	270	NC_010438
PEDV/CV777	Id	67	99-137	297	NC_003436
PEDV/LZC	Id	67	99-179	297	EF185992
BtCoV-512/2005	Id	70	97-135	294	NC_009657
IBV/Beaudette	IIIa	57	131-166	529	NC_001451
IBV/Beaudette-p65	IIIa	57	131-166	529	DQ001339
IBV/Peafowl	IIIa	57	131-166	529	AY641576
TCoV/MG10	IIIa	57	131-163	529	NC_010800
CoV-SW1	IIIb	72	-	524	NC_010646
BCoV	IIa	65	100-126	211	NC_003045
HEV	IIa	65	100-126	211	NC_007732
HCoV-OC43	IIa	65	100-126	211	NC_005147
HCoV-HKU1	IIa	66	98-118	206	NC_006577
MHV-A59	IIa	66	99-125	210	NC_001846
SARS-CoV/Tor2	IIb	67	104-136	265	NC_004718
SARS-CoV/TJF	IIb	66	103-135	264	AY654624
BtSCoV-Rp3	IIb	67	104-136	265	NC_009693
BtSCoV-HKU3-1	IIb	65	102-134	262	NC_009694
BtSCoV-279/2005	IIb	65	104-136	262	DQ648857
BtSCoV-Rm1	IIb	66	104-133	261	NC_009696
BtSCoV-273/2005	IIb	67	104-133	261	DQ648856
BtCoV-HKU9-4	IId	71	-	229	EF065516
BtCoV-HKU9-2	IId	69	-	228	EF065514
BtCoV-HKU9-3	IId	70	-	229	EF065515
BtCoV-HKU9-1	IId	71	-	229	NC_009021
BtCoV-HKU4-1	IIc	63	140-163	267	NC_009019
BtCoV-HKU4-4	IIc	63	140-163	267	EF065508
BtCoV-133/2005	IIc	56	133-156	260	NC_008315
BtCoV-HKU5-1	IIc	61	141-164	261	NC_009020
BtCoV-HKU5-2	IIc	61	140-163	260	EF065510
BuCoV-HKU11/796	IIIc	72	-	607	NC_011548
ThCoV-HKU12/600	IIIc	65	-	592	NC_011549
MuCoV-HKU13/3514	IIId	64	-	595	NC_011550

Fig. 1. Clustering and general features of the 5′ 420 nucleotides of CoVs. The tree is based on a multiple sequence alignment using ClustalW2 at the European Bioinformatics Institute webserver. The phylogenetic group, the start of core TRS-L, the region of upstream ORF (uORF), the start of ORF1ab, and GenBank accession number of each CoV are listed.

basically resembled the current grouping system for CoVs (Fig. 1), though group I CoVs may be further subdivided into 4 subgroups, groups Ia to Id, according to their relatively large phylogenetic distances (Fig. 1). Sequence comparison further showed conserved and unique features for each CoV group, including: (i) the relative location of the core sequence of the leader transcription-regulating sequence (L-TRS) is quite conserved in all CoVs, except for the one in group Ia CoVs which has a rather long leader sequence upstream of the core TRS; (ii) the potentially translatable short ORF upstream of the genomic ORF1ab, the uORF, is present in most CoVs except for group IId, IIIb, IIIc, and IIId CoVs; (iii) the 5′ UTR in group III CoVs is substantially longer than that in group I and II CoVs, while group IIa CoVs have an exclusively short 5′ UTR (Fig. 1). It has to be noted that in order to obtain a higher threshold of the phylogenetic distance, strains with the highest sequence variation were used for analysis (selected from the genomic sequences of all CoVs available in GenBank). This made it more promising if homology was found within a cluster. To further examine if particular features found in the RNA sequence in each group are relevant to specific organization of the 5′ *cis*-acting elements, we globally predicted the secondary structures of the CoV 5′ UTRs, predominantly using computational calculations at the *mfold* webserver (Zuker, 2003). We have identified several conserved stem–loop (SL) structures in this region, some of which are organized in a group-specific manner (see Figs. 2, 3, and 4).

Fig. 2. The structural–phylogenetic analysis of the 5′-proximal sequences in group I CoVs. The predicted secondary structures of the 5′-proximal sequence of (A) group Ia TGEV-purdue, (B) group Ib HCoV-229E-inf-1, (C) group Ic PEDV-CV777, and (D) group Id BtCoV-1A coronaviruses are shown. Nucleotide variations located in the conserved elements in the other representative CoVs of each subgroup are indicated. The start codon of the ORF1ab is boxed, the core sequence of the transcription-regulating leader (TRS-L CS) is bracketed, and the length of the sequence insertion in SL5 is indicated.


A) Group Ia
TGEV purdue
a:TGEV-TS
b:FIPV-79-1146
c:FCoV-Black
d:FCoV-C1Je
(106 nt)
TRS-L CS
5'
3'
5 52 53 67 72 83 91 149 154 327 339 350 371 436 441 518
SL1
SL2
SL2.1
SL4
SL5
B) Group Ib
HCoV-229E-inf-1
a:BtCoV-HKU2
b:HCoV-NL63-AMS1
c:HCoV-NL63-AMS496
(100 nt)
TRS-L CS
5'
3'
4 39 42 57 66 139 140 302 312 329 350 419
SL1
SL2
SL4
SL5

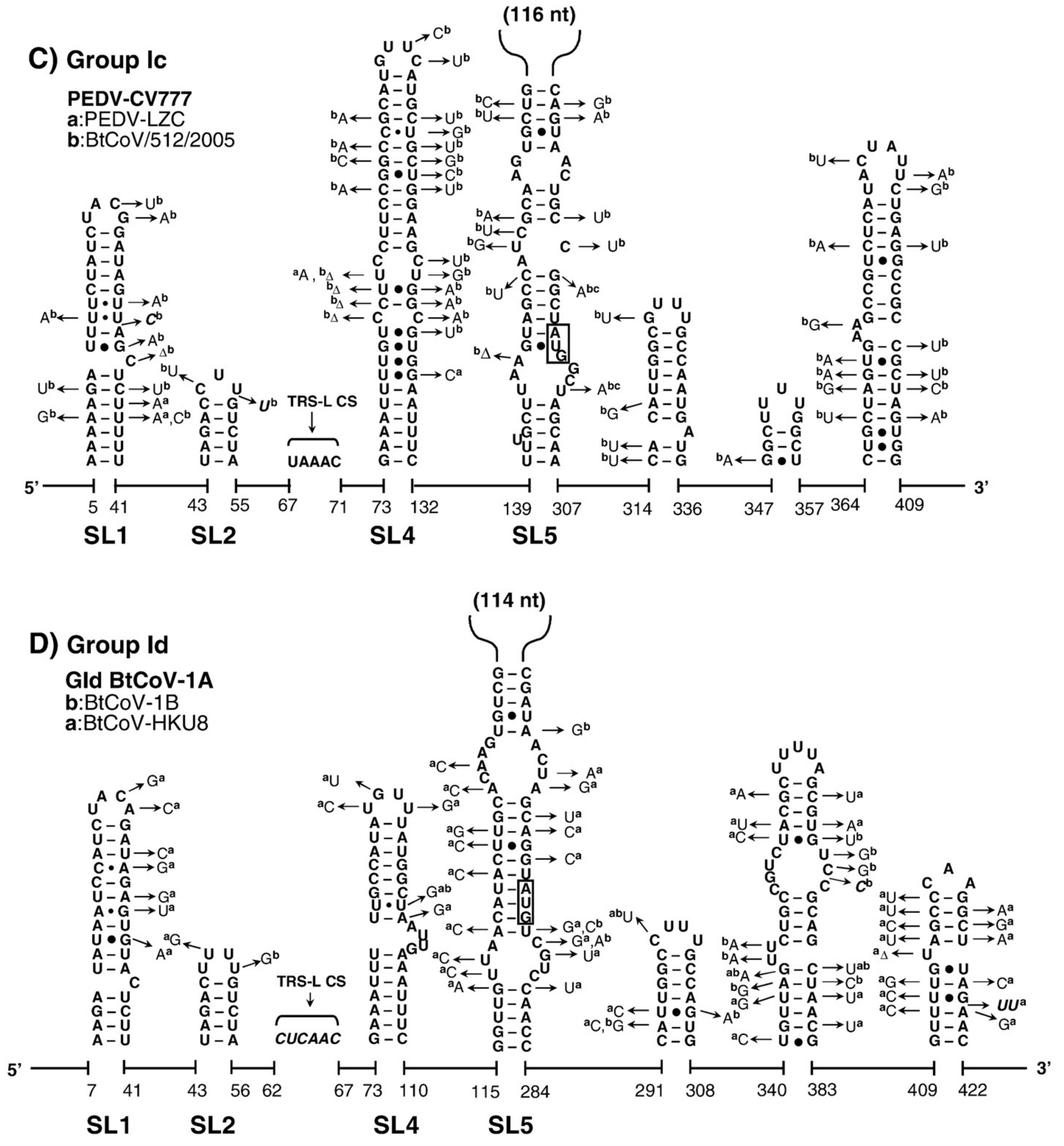


Fig. 2 (*continued*).

The universal presence of SL1 and SL2 in CoV 5′ UTR

The very 5′ nts of CoV RNAs fold into a hairpin of low thermodynamic stability, SL1, which is supported by many co-variations (Figs. 2–4), particularly in group IIa and IIIc CoVs. The loop sequences are not strongly conserved although a YRYR tetra-loop seems to be preferred in most SL1s. A general feature of SL1 is the presence of mismatches, bulges (*e.g.* in group I and II CoV RNAs) and a high number of A–U and U–A base pairs (bps) (*e.g.* in group IIIa, b, and d CoV RNAs). Recent data by Li et al. (2008) suggest that the low thermodynamic stability of SL1 is important for the replication of MHV.

Another conserved hairpin is SL2 which consists of a 5-bp stem and a highly conserved loop sequence, 5′-CUUGY-3′, which has an important role in MHV replication (Liu et al., 2007), though the motif is less conserved in SL2 of group I and III CoVs (Figs. 2 and 4). Downstream of SL2, an additional hairpin, SL2.1, with the stable UUCG tetra-loop, was predicted in group Ia CoVs. Interestingly, the CUUGY loop was recently shown to adopt the YNMG-type of tetra-loop-folds (Liu et al., 2009).

The diversity of SL3 and SL4 in CoVs

Previously, the core L-TRS in CoVs has either been proposed to be non-structured (Stirrups et al., 2000; Wang and Zhang, 2000) or to form a hairpin structure (Shieh et al., 1987; Chang et al., 1996). We found that the core L-TRS and the adjacent sequence may fold into SL3 in some CoVs, *e.g.* the group II *Bovine coronavirus* (BCoV), SARS-CoV and *Bat coronavirus HKU4* (BatCoV-HKU4), and the group III *coronavirus SW1* (CoV-SW1), *Bulbul coronavirus HKU11* (BuCoV-HKU11), and *Munia coronavirus HKU13* (MuCoV-HKU13) (Figs. 3 and 4). However, the sequence variations found in group IIa CoVs are partially in conflict with the lower part of SL3, while in other CoVs there are no co-variations to support the formation of SL3. Thus, the CoV SL3 may not structurally resemble the L-TRS Hairpin (LTH) found

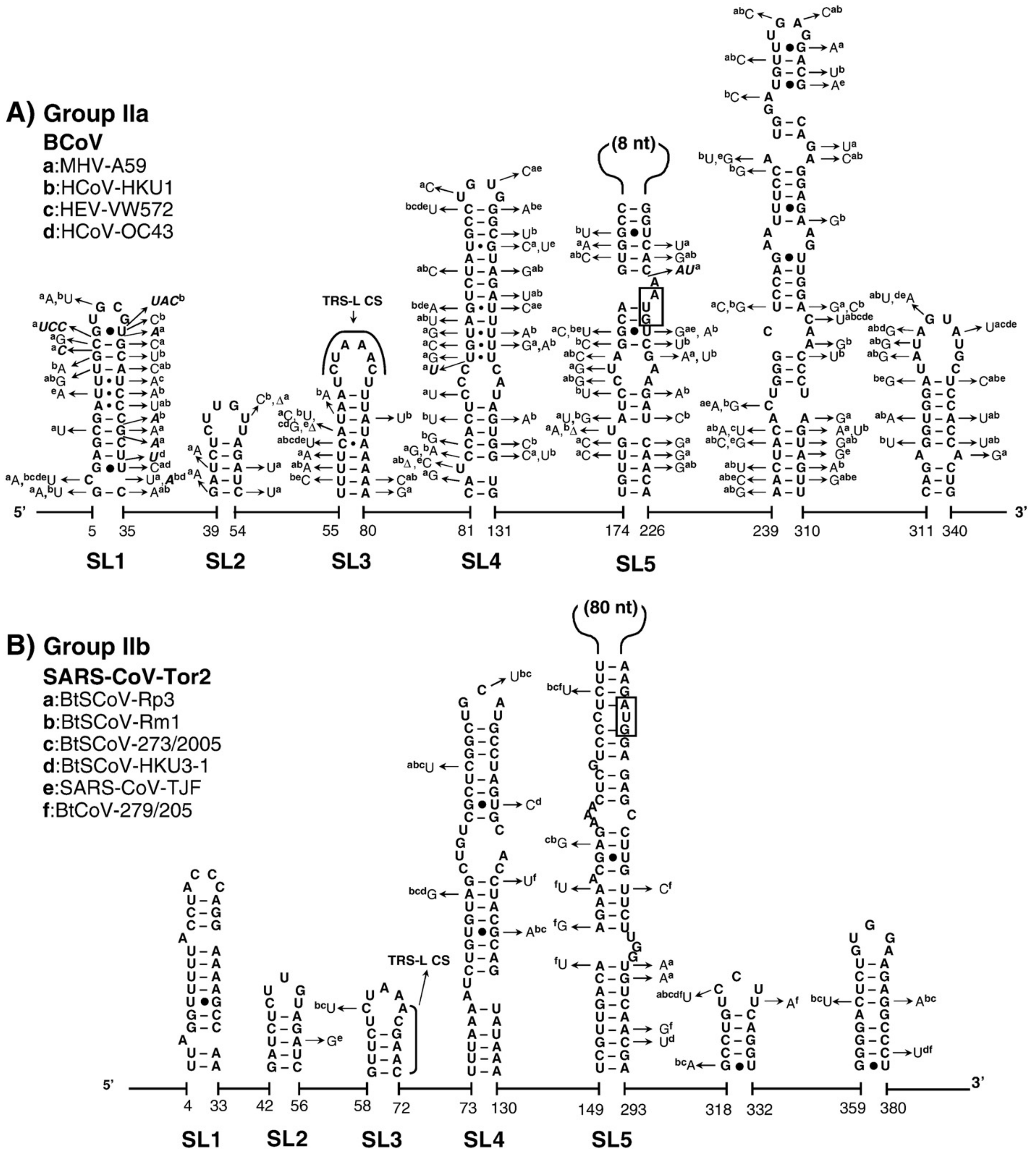


Fig. 3. The structural–phylogenetic analysis of the 5′-proximal sequences in group II CoVs. The predicted secondary structures of the 5′-proximal sequence of (A) group IIa BCoV, (B) group IIb SARS-CoV-Tor2, (C) group IIc BtCoV-HKU5-1, and (D) group IId BtCoV-HKU9-1 are shown. For details see Fig. 2.

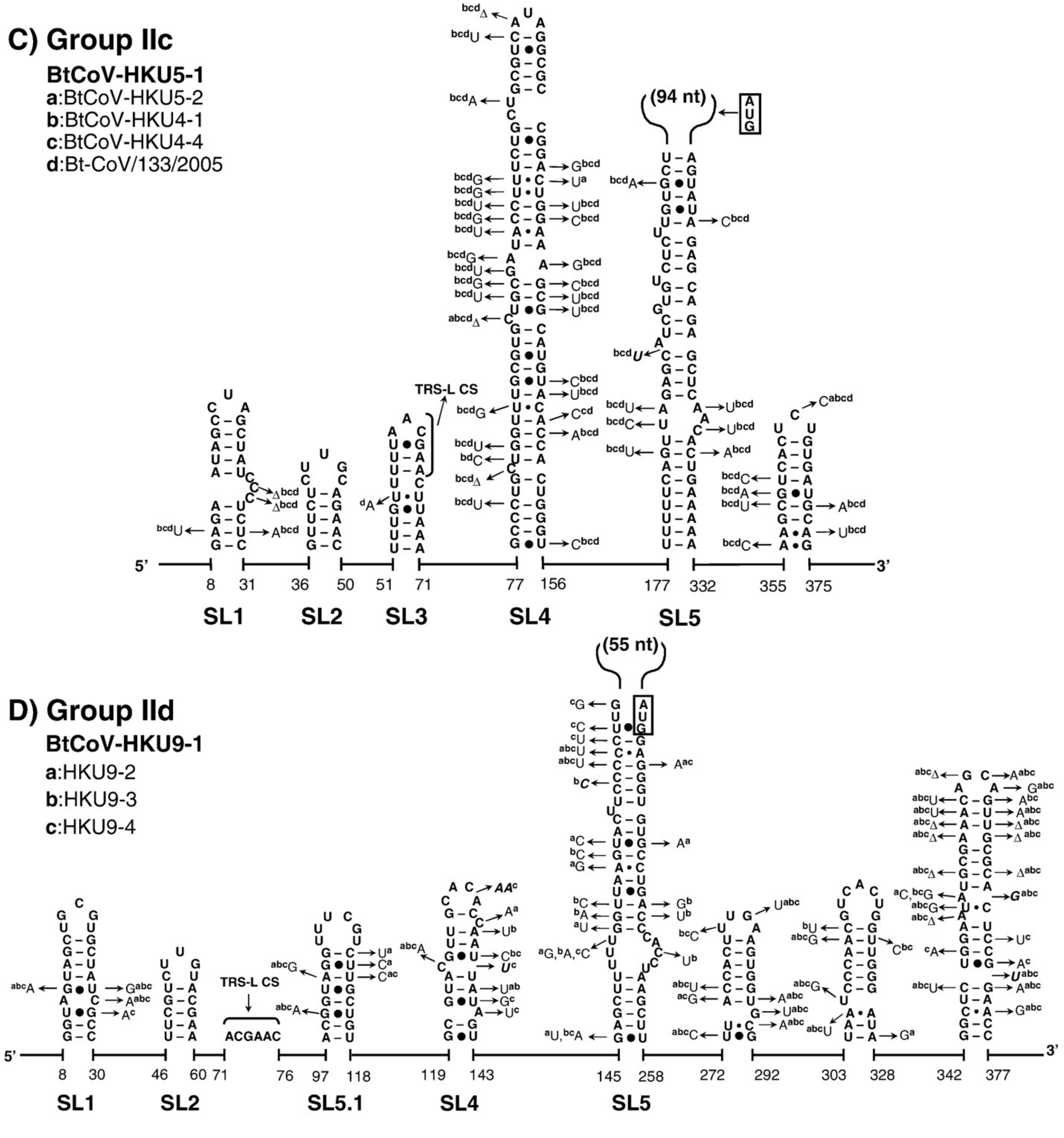


Fig. 3 (*continued*).

in the related arterivirus, the *Equine arteritis virus* (EAV), which directs discontinuous transcription (van den Born et al., 2004, 2005). In some other CoVs, *e.g.* TGEV and the *Human coronavirus-229E* (HCoV-229E), the core L-TRS was predicted to participate in the stem of SL4 (Figs. 2A and B), although sequence variations found in group Ib CoVs do not strongly support the involvement of the core L-TRS in the SL4 stem (Fig. 2B). All in all, based on the structural–phylogenetic survey, it can be concluded that the core L-TRS and the flanking sequences are poorly structured in CoVs.

Downstream of the L-TRS, a long hairpin, SL4, was predicted for all CoVs (Figs. 2, 3, and 4). The presence of a large number of co-variations seems to support the existence of SL4 strongly, particularly the upper half of this structure. Raman et al. (2003) have shown that the structural integrity, in positive or negative strands or both, of the upper part of SL4 (the SL-III in their study) is important for replication of BCoV DI RNA. We also found that the uORF predominantly terminates within the SL4 (data not shown), even for those uORFs that are in-frame with the downstream ORF1ab (Fig. 1).

There has no direct evidence for the translation of uORF in CoV infected cells, although Raman et al. (2003) have suggested a positive correlation between maintenance of the uORF and maximal BCoV DI RNA accumulation. They have also shown that a DI RNA in which this

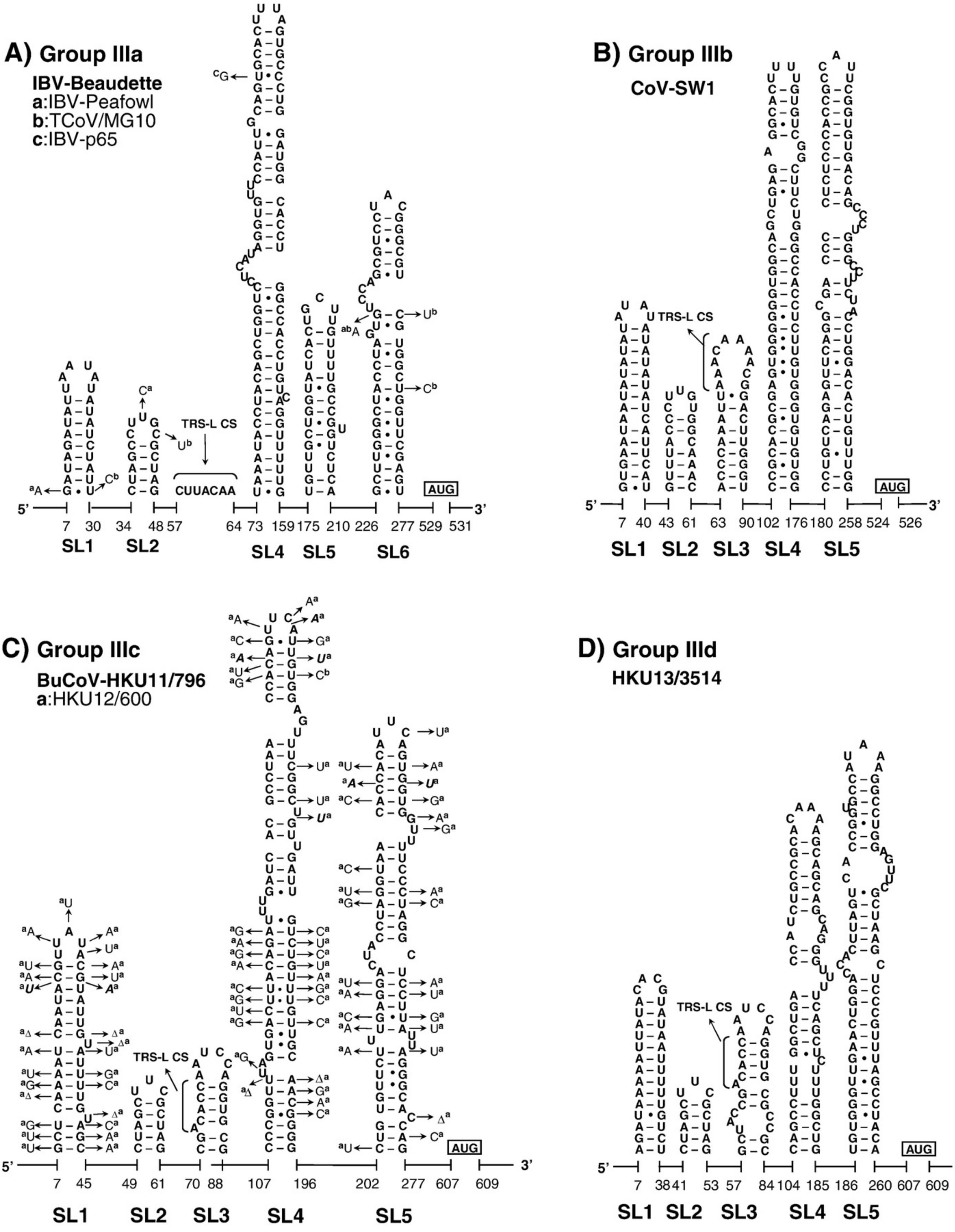
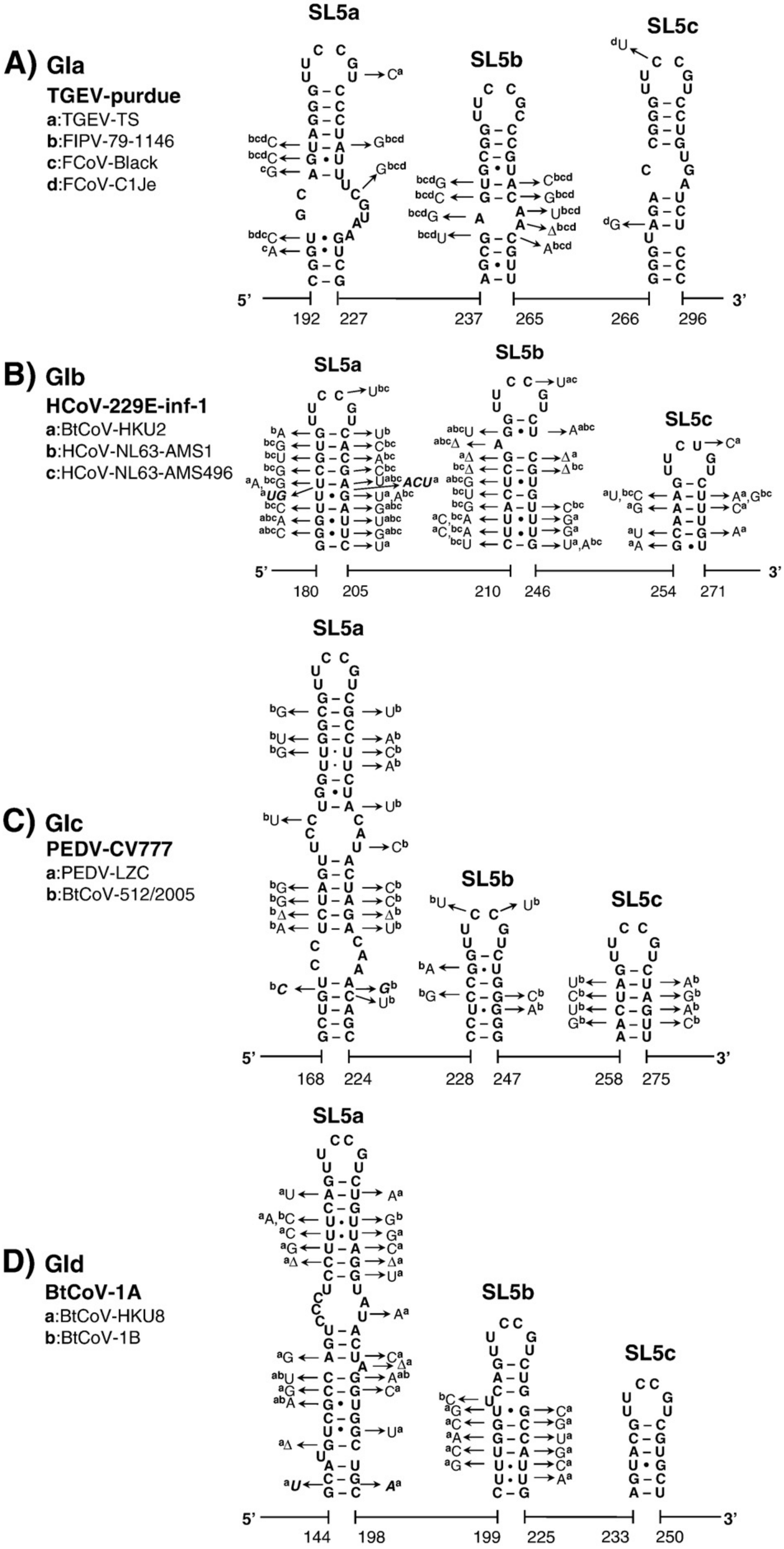


Fig. 4. The structural–phylogenetic analysis of the 5′-proximal sequences in group III CoVs. The predicted secondary structures of the 5′-proximal sequence of (A) group IIIa IBV-Beaudette, (B) group IIIb CoV-SW1, (C) group IIIc BuCoV-HKU11/796, and (D) group IIId HKU13/3514 are shown. For further details see Fig. 2.


A) GIa
TGEV-purdue
a:TGEV-TS
b:FIPV-79-1146
c:FCoV-Black
d:FCoV-C1Je
SL5a
SL5b
SL5c
5'
3'
192
227
237
265
266
296
B) GIb
HCoV-229E-inf-1
a:BtCoV-HKU2
b:HCoV-NL63-AMS1
c:HCoV-NL63-AMS496
SL5a
SL5b
SL5c
5'
3'
180
205
210
246
254
271
C) GIc
PEDV-CV777
a:PEDV-LZC
b:BtCoV-512/2005
SL5a
SL5b
SL5c
5'
3'
168
224
228
247
258
275
D) GId
BtCoV-1A
a:BtCoV-HKU8
b:BtCoV-1B
SL5a
SL5b
SL5c
5'
3'
144
198
199
225
233
250

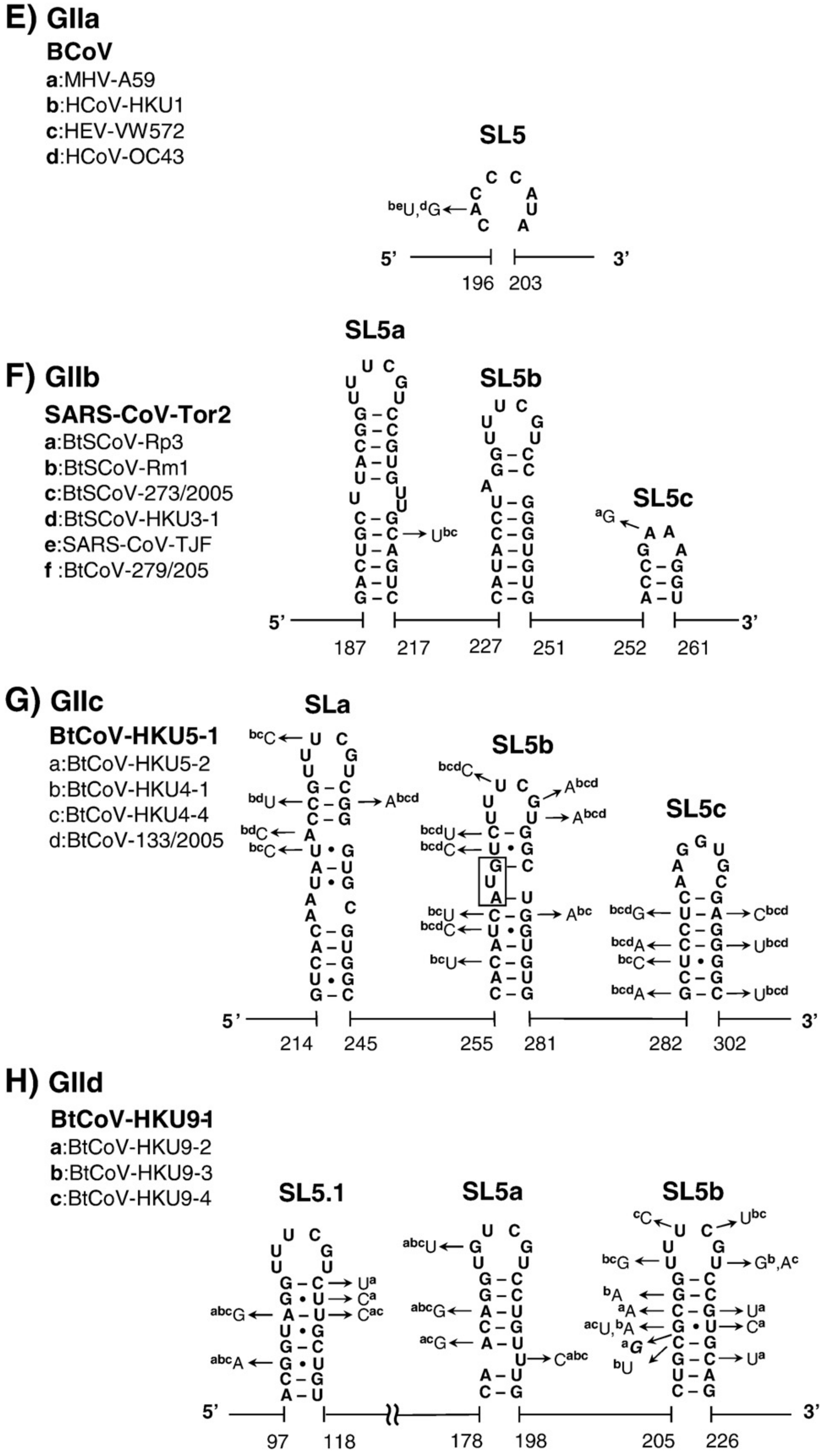
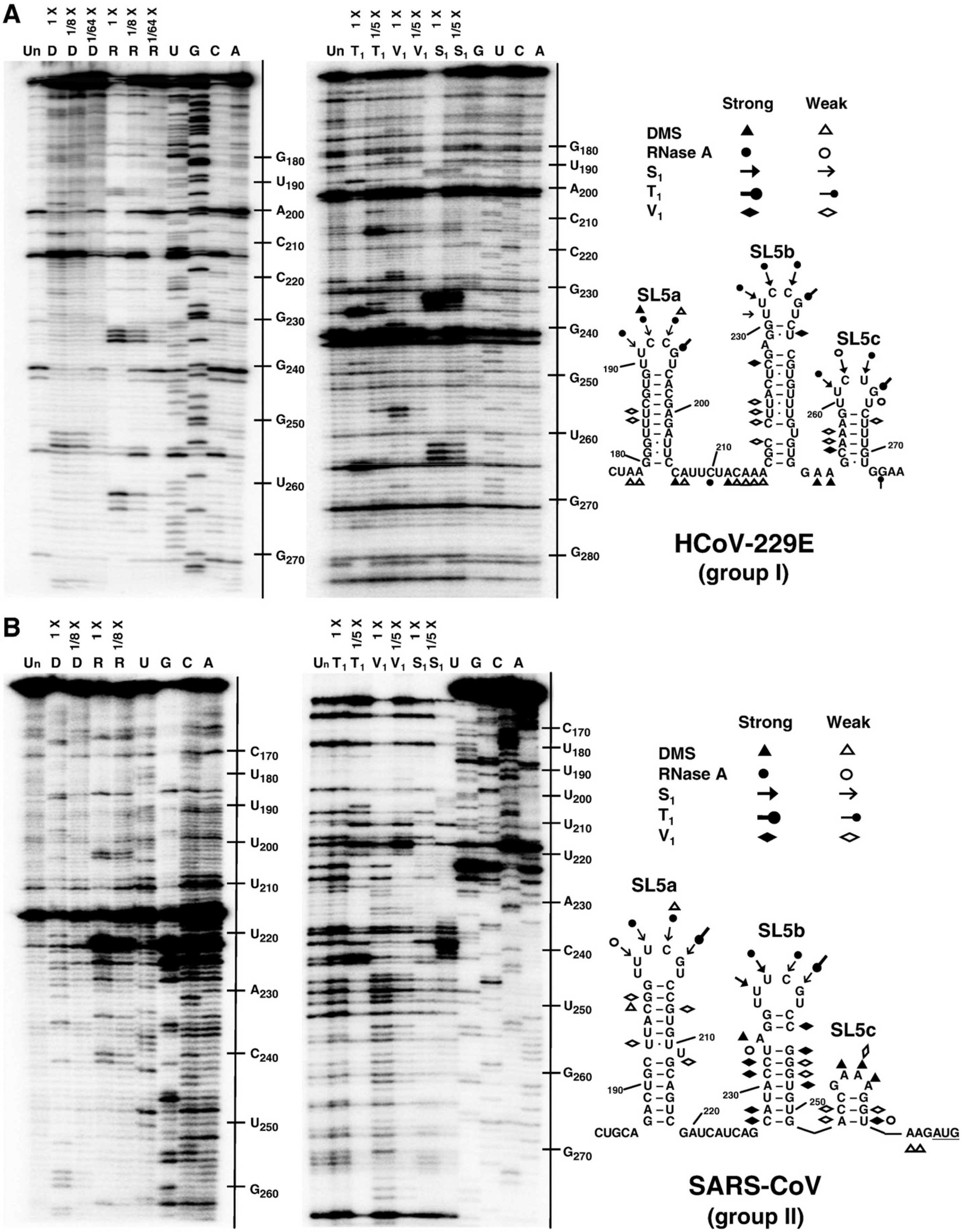


Fig. 5. The substructural hairpins of SL5 in group I and II CoVs. The secondary structure of the SL5 substructural hairpins, SL5a–c, in (A) group Ia TGEV-purdue, (B) group Ib HCoV-229E-inf-1, (C) group Ic PEDV-CV777, (D) group Id BtCoV-1A, (E) group IIa BCoV, (F) group IIb SARS-CoV-Tor2, (G) group IIc BtCoV-HKU5-1, and (H) group IId BtCoV-HKU9-1 are shown. The start codon of the BtCoV-HKU5-1 ORF1ab is located in SL5b as indicated. SL5.1 which is located upstream of SL5 in BtCoV-HKU9-1 also contains the conserved UUUCGU motif.


A
1 X
1/8 X
1/64 X
1 X
1/8 X
1/64 X
Un D D D R R R U G C A
G180
U190
A200
C210
C220
G230
G240
G250
U260
G270
1 X
1/5 X
1 X
1/5 X
1 X
1/5 X
Un T1 T1 V1 V1 S1 S1 G U C A
G180
U190
A200
C210
C220
G230
G240
G250
U260
G270
G280
Strong
Weak
DMS ▲ △
RNase A ● ○
S1
T1
V1 ◆ ◇
SL5a
SL5b
SL5c
180
190
200
210
230
260
270
CUAA
CAUUCUACAAA
GAA
GGAA
HCoV-229E
(group I)
B
1 X
1/8 X
1 X
1/8 X
Un D D R R U G C A
C170
U180
U190
U200
U210
U220
A230
C240
U250
G260
1 X
1/5 X
1 X
1/5 X
1 X
1/5 X
Un T1 T1 V1 V1 S1 S1 U G C A
C170
U180
U190
U200
U210
U220
A230
C240
U250
G260
G270
Strong
Weak
DMS ▲ △
RNase A ● ○
S1
T1
V1 ◆ ◇
SL5a
SL5b
SL5c
190
210
220
230
250
CUGCA
GAUCAUCAG
AAGAUG
SARS-CoV
(group II)

uORF was replaced by a totally unrelated uORF could be replicated. Our phylogenetic analysis showed that the sequence variations located in SL4, which were found to maintain the integrity of the RNA secondary structure, are not always silent at the amino acid level (data not shown). Although features of uORFs seem to be conserved and group-specific (Fig. 1), the necessity of translation of this ORF needs to be determined in the future to understand why certain groups of CoVs do need uORF for their propagation and others do not.

We noticed that the sequence of SL4 is included in the hotspot of the 5′-proximal genomic acceptor (Wu et al., 2006), suggesting that SL4 may play a role in directing the subgenomic RNA synthesis and thereby compensates for the absence of a structured L-TRS hairpin (see above).

Features of the inserted sequence in SL5 reflect the lineage of CoVs

A fifth structural element, SL5, was predicted downstream of SL4 in all CoVs (Figs. 2–4). SL5 is a homologue of SL-IV of BCoV reported by Brian and coworkers (Raman and Brian, 2005; Brown et al., 2007) and is supported by co-variations in almost all CoV groups with the exception of group Ia, and IIIa, b, and d CoVs, where sequence variation is low. Compared to group IIa and III CoVs, the other CoVs have sequence insertions in the top of SL5, which are about 110-nt long in group I CoVs and between 55 and 94 nt in group IIb, c, and d CoVs (Figs. 2, 3, and 4). Secondary structure predictions of these inserts revealed hairpins displaying the conserved 5′-UUYCGU-3′ loop motif (Fig. 5). We note that some of these hairpins resemble the predicted structures for four group I CoVs and SARS-CoV reported by Raman and Brian (2005), which were proposed to be homologues of BCoV SL5 (SL-IV in their report). Nevertheless, our comprehensive structural–phylogenetic analysis indicates that these conserved structural motifs are not SL5 homologues as such but are substructural hairpins within SL5 (Figs. 2, 3, and 5).

In group I CoVs, a large number of co-variations, particularly in group Ib CoVs, was observed, supporting the existence of these substructural hairpins at the top of SL5 (Fig. 5). We noticed that 4 different patterns of the SL5 substructural hairpins were found in group I CoVs. This finding supports the idea that group I CoVs may be clustered into 4 subgroups, groups Ia to d. Nonetheless, the structural homology of SL5 within the lineage of the group I CoVs is still higher than that of the group II CoVs; three hairpins, SL5a, b, and c, with mainly the conserved 5′-UUCCGU-3′ loop sequence, were found in all group I CoVs. This is in agreement with the shorter phylogenetic distances found between each subgroup (group Ia–d) in group I CoVs compared to group II CoVs, which feature more diverse sequence insertions, in terms of length, the presence of 5′-UUYCGU-3′ motifs, and secondary structure. The greater structural variation in SL5 of group II CoVs is as follows: (i) the substructural hairpins are replaced by an 8-nt sequence in group IIa CoVs (Fig. 5E); (ii) one of the three substructural hairpins in SL5, SL5c, contains a GNRA tetra-loop sequence (group IIb) or a non-conserved hepta-loop sequence (group IIc) but not the UUYCGU motif (Figs. 5F and G); (iii) only two substructural hairpins are folded on top of SL5 in group IId CoVs, yet an additional conserved UUCCGU motif is present in SL5.1 located further upstream, in between the L-TRS and SL4 (Figs. 3D and 5H). Thus, the pattern of the SL5 substructures is strongly related to the lineage and the phylogenetic distance of the group I and II CoVs.

Similar hairpins with a conserved loop motif could not be identified in group III CoVs (Fig. 4). Here, SL5 has a rod-like shape as in group IIa. Also in the remainder of the 5′ UTR of group III CoVs no hairpins could be identified that featured a UUYCGU sequence or another motif.

Structure probing of the SL5 substructure in HCoV-229E and SARS-CoV

To verify the secondary structures of the proposed substructural hairpins in group I and II CoVs, the corresponding RNA transcripts of HCoV-229E and SARS-CoV were subjected to enzymatic and chemical structure probing (see Materials and methods). Clearly, the single-stranded 5′-UUUCGU-3′-hexa-loop sequences in HCoV-229E SL5a, SL5b, and SL5c can be recognized by the single-strand specific probes, DMS, RNase A, S1 nuclease and/or RNase T1 (Fig. 6A), suggesting that these nucleotides are unpaired. The presence of RNase V1 cuts, an enzyme that cuts double-stranded RNA, in the predicted stem regions is also in agreement with the model. Probing results of SARS-CoV were also in agreement with the existence of SL5a, b, and c (Fig. 6B).

Notably, the U:U mismatches located in the stems of these substructures seem to form non-canonical base pairs since RNAse V1 recognized U222 and U221 in HCoV-229E SL5b, as well as U193 in SARS-CoV SL5a. In fact, several (tandem) U:U mismatches were identified in the SL5 substructural hairpins, *e.g.* the SL5a in the group Ic *Porcine epidemic diarrhea virus* (PEDV) (Fig. 5C) and the group Id *Bat coronavirus 1A* (BtCoV-1A) (Fig. 5D), as well as in other 5′ *cis*-acting elements, *e.g.* the MHV SL1 (Li et al., 2008). Interestingly, co-variations were frequently found at the positions of these tandem U:U mismatches, *e.g.* SL4 (Figs. 2B, 3C, 4C) and SL5a–c (Figs. 5B, C, D, and H). This suggests the formation of (tandem) U:U base pairs similar to what has been reported for the 5′-CU-3′/5′-UU-3′ non-canonical base pairs found in the Y stem of polio-like enterovirus 3′ UTRs (Lescrinier et al., 2003).

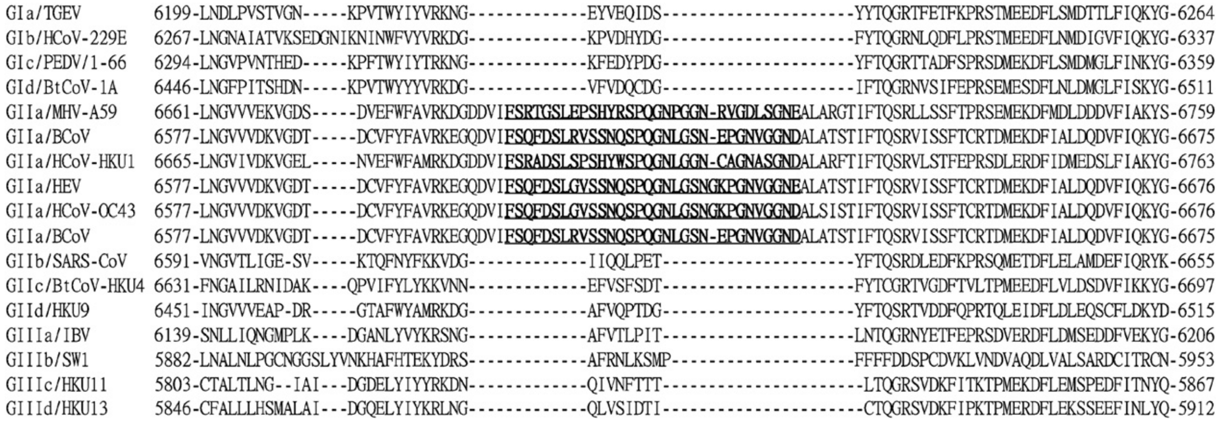
Are the SL5 substructural hairpins the counterparts of the group IIa packaging signal?

It has been generally found that a strong packaging signal (PS) or encapsidation signal, which directs specific packaging or encapsidation of genomic RNA, usually encompasses repetitions of conserved (structural) motifs (Hellendoorn et al., 1996, Chen et al., 2007). This leads us to propose that the SL5 substructures bearing the highly conserved UUYCGU repeats function as genomic PS for group I and II CoVs, including SARS-CoV.

Studies of the genomic PSs in CoVs have been mainly focused on group IIa CoVs in the past, e.g. MHV and BCoV (Fosmire et al., 1992; Makino et al., 1990; van der Most et al., 1991; Woo et al., 1997; Chen et al., 2007; Cologna and Hogue, 2000). For other groups of CoVs, e.g. SARS-CoV, the identification of a putative PS has been reported by Hsieh et al. (2005). This PS was thought to be a homologue of the MHV PS located in the corresponding region near the 3′ end of ORF1ab. However, it has to be noted that the specificity of the proposed SARS-CoV PS to direct RNA packaging was not determined in their study, and the predicted secondary structure of their "homologue of MHV PS" lacks the conserved features of the MHV PS structure reported by Chen et al. (2007). Also we doubt the possibility of identifying a MHV-like PS in the "corresponding region" of SARS-CoV genomic RNA because an alignment of nsp15 sequences clearly shows that the sequence corresponding to the MHV PS is absent in SARS and other non-group IIa CoVs (Fig. 7).

Interestingly, the presence or absence of the region corresponding to the group IIa PS may not interfere with the function of nsp15 as the functional domains remain intact in both MHV and SARS-CoV nsp15 (Joseph et al., 2007). There seems to be however a strong correlation

Fig. 6. Structure probing of the inserted sequences in SL5 of group Ib HCoV-229E and group IIb SARS-CoV-Tor2. The secondary structures of the SL5 substructural hairpins of (A) the HCoV-229E and (B) the SARS-CoV are analyzed by enzymatic and chemical structure probing. Annotation of the denaturing electrophoresis: Un, untreated; D, DMS treated; R, RNase A treated; T_1, RNase T_1 treated; V_1, RNase V_1 treated; S_1, S_1 nuclease treated; G, U, C and A, the RNA sequencing ladder.



```
GIa/TGEV         6199-LNDLPVSTVGN-----KPVTWYIYVRKNG-------------EYVEQIDS---------------------YYTQGRTFETFKPRSTMEEDFLSMDTTLFIQKYG-6264
GIb/HCoV-229E    6267-LNGNAIATVKSEDGNIKNINWFVYVRKDG-------------KPVDHYDG---------------------FYTQGRNLQDFLPRSTMEEDFLNMDIGVFIQKYG-6337
GIc/PEDV/1-66    6294-LNGVPVNTHED-----KPFTWYIYTRKNG-------------KFEDYPDG---------------------YFTQGRTTADFSPRSDMEKDFLSMDMGLFINKYG-6359
GId/BtCoV-1A     6446-LNGFPITSHDN-----KPVTWYYYVRKDG-------------VFVDQCDG---------------------IFTQGRNVSIFEPRSEMESDFLNLDMGLFISKYG-6511
GIIa/MHV-A59     6661-LNGVVVEKVGDS-----DVEFWFAVRKDGDDVIESRTGSLEPSHYRSPQGNPGGN-RVGDLSGNEALARGTIFTQSRLLSSFTPRSEMEKDFMDLDDDVFIAKYS-6759
GIIa/BCoV        6577-LNGVVVDKVGDT-----DCVFYFAVRKEGQDVIESQFDSLRVSSNQSPQGNLGSN-EPGNVGGNDALATSTIFTQSRVISSFTCRTDMEKDFIALDQDVFIQKYG-6675
GIIa/HCoV-HKU1   6665-LNGVIVDKVGEL-----NVEFWFAMRKDGDDVIESRADSLSPSHYWSPQGNLGGN-CAGNASGNDALARFTIFTQSRVLSTFEPRSDLERDFIDMEDSLFIAKYG-6763
GIIa/HEV         6577-LNGVVVDKVGDT-----DCVFYFAVRKEGQDVIESQFDSLGVSSNQSPQGNLGSNGKPGNVGGNEALATSTIFTQSRVISSFTCRTDMEKDFIALDQDVFIQKYG-6676
GIIa/HCoV-OC43   6577-LNGVVVDKVGDT-----DCVFYFAVRKEGQDVIESQFDSLGVSSNQSPQGNLGSNGKPGNVGGNDALSISTIFTQSRVISSFTCRTDMEKDFIALDQDVFIQKYG-6676
GIIa/BCoV        6577-LNGVVVDKVGDT-----DCVFYFAVRKEGQDVIESQFDSLRVSSNQSPQGNLGSN-EPGNVGGNDALATSTIFTQSRVISSFTCRTDMEKDFIALDQDVFIQKYG-6675
GIIb/SARS-CoV    6591-VNGVTLIGE-SV-----KTQFNYFKKVDG-------------IIQQLPET---------------------YFTQSRDLEDFKPRSQMETDFLELAMDEFIQRYK-6655
GIIc/BtCoV-HKU4  6631-FNGAILRNIDAK----QPVIFYLYKKVNN-------------EFVSFSDT---------------------FYTCGRTVGDFTVLTPMEEDFLVLDSDVFIKKYG-6697
GIId/HKU9        6451-INGVVVEAP-DR-----GTAFWYAMRKDG-------------AFVQPTDG---------------------YFTQSRTVDDFQPRTQLEIDFLDLEQSCFLDKYD-6515
GIIIa/IBV        6139-SNLLIQNGMPLK----DGANLYVYKRSNG-------------AFVTLPIT---------------------LNTQGRNYETFEPRSDVERDFLDMSEDDFVEKYG-6206
GIIIb/SW1        5882-LNALNLPGCNGGSLYVNKHAFHTEKYDRS-------------AFRNLKSMP--------------------FFFFDDSPCDVKLVNDVAQDLVALSARDCITRCN-5953
GIIIc/HKU11      5803-CTALTLNG--IAI---DGDELYIYYRKDN-------------QIVNFTTT----------------------LTQGRSVDKFITKTPMEKDFLEMSPEDFITNYQ-5867
GIIId/HKU13      5846-CFALLLHSMALAI---DGQELYIYKRLNG-------------QLVSIDTI----------------------CTQGRSVDKFIPKTPMERDFLEKSSEEFINLYQ-5912
```

Fig. 7. Multiple alignment of the CoV nsp15 sequence corresponding to the group IIa packaging signal. The amino acid sequences of the group IIa CoV nsp15 are aligned with the sequences of other CoV groups, showing the underlined sequence insertion of the packaging signal corresponding region in group IIa CoVs.

between the lack of a MHV PS-corresponding region and the presence of SL5 substructures and vice versa (Fig. 5). This correlation strongly suggests that the SL5 substructural hairpins located in the 5′ UTR are the counterparts of the genomic PS present in group IIa CoVs, and presumably the UUCCGU structural repeats (Fig. 5A) are responsible for the packaging activity reported by Escors et al. (2003) for the first 649 nts of TGEV genomic RNA.

Conclusions

The diversity of the genomic RNA sequence provides a wealth of structural and phylogenetic information on the lineage of CoVs and improves our understanding of the evolution of the 5′ *cis*-acting elements. We have shown that the pattern of these *cis*-acting elements in the 5′ UTR is highly related to the phylogenetic distance based on the viral protein sequences, suggesting that the viral proteins and the RNA sequence evolved simultaneously, possibly to maintain functional RNA–protein interactions.

The unique and conserved features of the 5′ UTR and SL5 highlight the role of RNA structure in the evolution of CoVs and may serve as a roadmap for further studies. Future experiments should also verify whether the conserved UUYCGU motifs in SL5 function as PS in group I and II CoVs by interacting with nucleocapsid and/or membrane proteins (Molenkamp and Spaan, 1997; Narayanan and Makino, 2001; Narayanan et al., 2003). The absence of these or other conserved motifs in the 5′ UTR of group III CoVs suggests that their PSs are located elsewhere in the genome. This possibility is currently being explored.

Materials and methods

Structural–phylogenetic analysis

Multiple alignment of all CoV 5′-proximal sequences available in GenBank was used to select coronaviruses with the highest sequence diversity. Sequences of the 5′ 420 nts of these variants were clustered by ClustalW2 on EBI webserver (Larkin et al., 2007). Secondary structures of this region were predicted by the Mfold webserver (Zuker, 2003). The alignment of CoV nsp15 was done by Kalign webserver (Lassmann and Sonnhammer, 2006) (Fig. 7).

Structure probing and primer extension

The RNA transcripts encompassing the entire HCoV-229E and SARS-CoV SL5 region (about 180 nt) were synthesized *in vitro* using Ribomax^{TM} RNA production system (Promega). The corresponding cDNA templates with an upstream T7 promoter were amplified by PCR using oligo-nucleotides 5′-TAATACGACTCACTATAGGGCATGCCTAGTGCACCTACGCAG-3′ (the T7 promoter sequence is underlined) and 5′-CAAACTGAGTTGGACGTGTG-3′ for SARS-CoV SL5 and oligonucleotides 5′-TAATACGACTCACTATAGGGTAATTGAAATTTCATTTGGG-3′ (the T7 promoter sequence is underlined) and 5′-GTGTGACACTTGCCGTAGC-3′ for HCoV-229E SL5. Purified RNA transcripts were subjected to chemical and enzymatic probing as described in Chen et al. (2007). In general, 0.001% dimethylsulfate (DMS), 1 pg Rnase A, 0.001 U RNase T1, 0.1 U RNase V1, and 0.8 U S1 nuclease were used for the probing reactions (1×), followed by serial dilutions with a factor 1/5 (1/5× and 1/25×) or 1/8 (1/8× and 1/64×). The primer extension was carried out with 0.01 µg of treated transcripts, 0.5 µl of a 0.1 mM concentration of the MHV1 primer, 1 µl of 5 mM dGAT, 1 µl of 25 µM dCTP, 0.1 µl of α-^{32}P-labeled dCTP (10 mCi/ml), 1 µl of 5× reverse transcriptase buffer, and 20 U of Moloney murine leukemia virus reverse transcriptase (Promega).

Acknowledgments

We acknowledge Eric Snijder for providing us with partial cDNA corresponding to the 5′-proximal region of HCoV-229E and SARS-CoV. This research was supported by a VIDI grant from the Netherlands Organisation for Scientific Research (NWO) awarded to R.C.L.O.

References

Brian, D.A., Baric, R.S., 2005. Coronavirus genome structure and replication. Curr. Top. Microbiol. Immunol. 287, 1–30.

Brown, C.G., Nixon, K.S., Senanayake SDChang, R.Y., Hofmann, M.A., Sethna, P.B., Brian, D.A., 2007. An RNA stem–loop within the bovine coronavirus nsp1 coding region is a *cis*-acting element in defective interfering RNA replication. J. Virol. 81, 7716–7724.

Chang, R.Y., Hofmann, M.A., Sethna, P.B., Brian, D.A., 1994. A *cis*-acting function for the coronavirus leader in defective interfering RNA replication. J. Virol. 68 (12), 8223–8231.

Chang, R.Y., Krishnan, R., Brian, D.A., 1996. The UCUAAAC promoter motif is not required for high-frequency leader recombination in bovine coronavirus defective interfering RNA. J. Virol. 70 (5), 2720–9272.

Chen, S.C., van den Born, E., van den Worm, S.H., Pleij, C.W., Snijder, E.J., Olsthoorn, R.C., 2007. New structure model for the packaging signal in the genome of group IIa coronaviruses. J. Virol. 81 (12), 6771–6774.

Cologna, R., Hogue, B.G., 2000. Identification of a bovine coronavirus packaging signal. J. Virol. 74 (1), 580–583.

Dong, B.Q., Liu, W., Fan, X.H., Vijaykrishna, D., Tang, X.C., Gao, F., Li, L.F., Li, G.J., Zhang, J.X., Yang, L.Q., Poon, L.L., Zhang, S.Y., Peiris, J.S., Smith, G.J., Chen, H., Guan, Y., 2007. Detection of a novel and highly divergent coronavirus from Asian leopard cats and Chinese ferret badgers in Southern China. J. Virol. 81 (13), 6920–6926.

Escors, D., Izeta, A., Capiscol, C., Enjuanes, L., 2003. Transmissible gastroenteritis coronavirus packaging signal is located at the 5′ end of the virus genome. J. Virol. 77 (14), 7890–7902.

Fosmire, J.A., Hwang, K., Makino, S., 1992. Identification and characterization of a coronavirus packaging signal. J. Virol. 66 (6), 3522–3530.

Gibbs, A.J., Gibbs, M.J., Armstrong, J.S., 2004. The phylogeny of SARS coronavirus. Arch. Virol. 149 (3), 21–624.

Goebel, S.J., Hsue, B., Dombrowski, T.F., Masters, P.S., 2004a. Characterization of the RNA components of a putative molecular switch in the 3′ untranslated region of the murine coronavirus genome. J. Virol. 78 (2), 669–682.

Goebel, S.J., Taylor, J., Masters, P.S., 2004b. The 3′ *cis*-acting genomic replication element of the severe acute respiratory syndrome coronavirus can function in the murine coronavirus genome. J. Virol. 78 (14), 7846–7851.

Goebel, S.J., Miller, T.B., Bennett, C.J., Bernard, K.A., Masters, P.S., 2007. A hypervariable region within the 3′ *cis*-acting element of the murine coronavirus genome is nonessential for RNA synthesis but affects pathogenesis. J. Virol. 81 (3), 1274–1287.

Guan, Y., Zheng, B.J., He, Y.Q., Liu, X.L., Zhuang, Z.X., Cheung, C.L., Luo, S.W., Li, P.H., Zhang, L.J., Guan, Y.J., Butt, K.M., Wong, K.L., Chan, K.W., Lim, W., Shortridge, K.F., Yuen, K.Y., Peiris, J.S., Poon, L.L., 2003. Isolation and characterization of viruses related to the SARS coronavirus from animals in southern China. Science 302 (5643), 276–278.

Hellendoorn, K., Michiels, P.J., Buitenhuis, R., Pleij, C.W.A., 1996. Protonatable hairpins are conserved in the 5′-untranslated region of tymovirus RNAs. Nucleic Acids Res. 24 (24), 4910–4917.

Hsieh, P.K., Chang, S.C., Huang, C.C., Lee, T.T., Hsiao, C.W., Kou, Y.H., Chen, I.Y., Chang, C.K., Huang, T.H., Chang, M.F., 2005. Assembly of severe acute respiratory syndrome coronavirus RNA packaging signal into virus-like particles is nucleocapsid dependent. J. Virol. 79 (22), 13848–13855.

Joseph, J.S., Saikatendu, K.S., Subramanian, V., Neuman, B.W., Buchmeier, M.J., Stevens, R.C., Kuhn, P., 2007. Crystal structure of a monomeric form of severe acute respiratory syndrome coronavirus endonuclease nsp15 suggests a role for hexamerization as an allosteric switch. J. Virol. 81 (12), 6700–6708.

Ksiazek, T.G., Erdman, D., Goldsmith, C.S., Zaki, S.R., Peret, T., Emery, S., Tong, S., Urbani, C., Comer, J.A., Lim, W., Rollin, P.E., Dowell, S.F., Ling, A.E., Humphrey, C.D., Shieh, W.J., Guarner, J., Paddock, C.D., Rota, P., Fields, B., DeRisi, J., Yang, J.Y., Cox, N., Hughes, J.M., LeDuc, J.W., Bellini, W.J., Anderson, L.J., SARS Working Group, 2003. A novel coronavirus associated with severe acute respiratory syndrome. N. Engl. J. Med. 348 (20), 1953–1966.

Lai, M.M., Cavanagh, D., 1997. The molecular biology of coronaviruses. Adv. Virus Res. 48, 1–100.

Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R., Thompson, J.D., Gibson, T.J., Higgins, D.G., 2007. ClustalW and ClustalX version 2. Bioinformatics 23 (21), 2947–2948.

Lassmann, T., Sonnhammer, E.L.L., 2006. Kalign, Kalignvu and Mumsa: web servers for multiple sequence alignment. Nucleic Acids Res. 34, W596–W599 Web Server issue.

Lescrinier, E.M., Tessari, M., van Kuppeveld, F.J., Melchers, W.J., Hilbers, C.W., Heus, H.A., 2003. Structure of the pyrimidine-rich internal loop in the poliovirus 3′-UTR: the importance of maintaining pseudo-2-fold symmetry in RNA helices containing two adjacent non-canonical base-pairs. J. Mol. Biol. 331 (4), 759–769.

Li, W., Shi, Z., Yu, M., Ren, W., Smith, C., Epstein, J.H., Wang, H., Crameri, G., Hu, Z., Zhang, H., Zhang, J., McEachern, J., Field, H., Daszak, P., Eaton, B.T., Zhang, S., Wang, L.F., 2005. Bats are natural reservoirs of SARS-like coronaviruses. Science 310 (5748), 676–679.

Li, L., Kang, H., Liu, P., Makkinje, N., Williamson, S.T., Leibowitz, J.L., Giedroc, D.P., 2008. Structural lability in stem–loop 1 drives a 5′ UTR–3′ UTR interaction in coronavirus replication. J. Mol. Biol. 377 (3), 790–803.

Liu, P., Li, L., Millership, J.J., Kang, H., Leibowitz, J.L., Giedroc, D.P., 2007. A U-turn motif-containing stem–loop in the coronavirus 5′ untranslated region plays a functional role in replication. RNA 13 (5), 763–780.

Liu, P., Li, L., Keane, S.C., Yang, D., Leibowitz, J.L., Giedroc, D.P., 2009. Mouse hepatitis virus stem–loop 2 adopts a uYNMG(U)a-like tetraloop structure that is highly functionally tolerant of base substitutions. J. Virol. 83 (23), 12084–12093.

Makino, S., Yokomori, K., Lai, M.M., 1990. Analysis of efficiently packaged defective interfering RNAs of murine coronavirus: localization of a possible RNA-packaging signal. J. Virol. 64 (12), 6045–6053.

Marra, M.A., Jones, S.J., Astell, C.R., Holt, R.A., Brooks-Wilson, A., Butterfield, Y.S., Khattra, J., Asano, J.K., Barber, S.A., Chan, S.Y., Cloutier, A., Coughlin, S.M., Freeman, D., Girn, N., Griffith, O.L., Leach, S.R., Mayo, M., McDonald, H., Montgomery, S.B., Pandoh, P.K., Petrescu, A.S., Robertson, A.G., Schein, J.E., Siddiqui, A., Smailus, D.E., Stott, J.M., Yang, G.S., Plummer, F., Andonov, A., Artsob, H., Bastien, N., Bernard, K., Booth, T.F., Bowness, D., Czub, M., Drebot, M., Fernando, L., Flick, R., Garbutt, M., Gray, M., Grolla, A., Jones, S., Feldmann, H., Meyers, A., Kabani, A., Li, Y., Normand, S., Stroher, U., Tipples, G.A., Tyler, S., Vogrig, R., Ward, D., Watson, B., Brunham, R.C., Krajden, M., Petric, M., Skowronski, D.M., Upton, C., Roper, R.L., 2003. The genome sequence of the SARS-associated coronavirus. Science 300 (5624), 1399–1404.

Mihindukulasuriya, K.A., Wu, G., St, Leger J., Nordhausen, R.W., Wang, D., 2008. Identification of a novel coronavirus from a beluga whale by using a panviral microarray. J. Virol. 82 (10), 5084–5088.

Molenkamp, R., Spaan, W.J., 1997. Identification of a specific interaction between the coronavirus mouse hepatitis virus A59 nucleocapsid protein and packaging signal. Virology 239 (1), 78–86.

Narayanan, K., Makino, S., 2001. Cooperation of an RNA packaging signal and a viral envelope protein in coronavirus RNA packaging. J. Virol. 75 (19), 9059–9067.

Narayanan, K., Chen, C.J., Maeda, J., Makino, S., 2003. Nucleocapsid-independent specific viral RNA packaging via viral envelope protein and viral RNA signal. J. Virol. 77 (5), 2922–2927.

Raman, S., Brian, D.A., 2005. Stem–loop IV in the 5′ untranslated region is a *cis*-acting element in bovine coronavirus defective interfering RNA replication. J. Virol. 79 (19), 12434–12446.

Raman, S., Bouma, P., Williams, G.D., Brian, D.A., 2003. Stem–loop III in the 5′ untranslated region is a *cis*-acting element in bovine coronavirus defective interfering RNA replication. J. Virol. 77 (12), 6720–6730.

Rota, P.A., Oberste, M.S., Monroe, S.S., Nix, W.A., Campagnoli, R., Icenogle, J.P., Peñaranda, S., Bankamp, B., Maher, K., Chen, M.H., Tong, S., Tamin, A., Lowe, L., Frace, M., DeRisi, J.L., Chen, Q., Wang, D., Erdman, D.D., Peret, T.C., Burns, C., Ksiazek, T.G., Rollin, P.E., Sanchez, A., Liffick, S., Holloway, B., Limor, J., McCaustland, K., Olsen-Rasmussen, M., Fouchier, R., Günther, S., Osterhaus, A.D., Drosten, C., Pallansch, M.A., Anderson, L.J., Bellini, W.J., 2003. Characterization of a novel coronavirus associated with severe acute respiratory syndrome. Science 300 (5624), 1394–1399.

Shieh, C.K., Soe, L.H., Makino, S., Chang, M.F., Stohlman, S.A., Lai, M.M., 1987. The 5′-end sequence of the murine coronavirus genome: implications for multiple fusion sites in leader-primed transcription. Virology 156 (2), 321–330.

Snijder, E.J., Bredenbeek, P.J., Dobbe, J.C., Thiel, V., Ziebuhr, J., Poon, L.L., Guan, Y., Rozanov, M., Spaan, W.J., Gorbalenya, A.E., 2003. Unique and conserved features of genome and proteome of SARS-coronavirus, an early split-off from the coronavirus group 2 lineage. J. Mol. Biol. 331 (5), 991–1004.

Stirrups, K., Shaw, K., Evans, S., Dalton, K., Cavanagh, D., Britton, P., 2000. Leader switching occurs during the rescue of defective RNAs by heterologous strains of the coronavirus infectious bronchitis virus. J. Gen. Virol. 81 (Pt 3), 791–801.

Tang, X.C., Zhang, J.X., Zhang, S.Y., Wang, P., Fan, X.H., Li, L.F., Li, G., Dong, B.Q., Liu, W., Cheung, C.L., Xu, K.M., Song, W.J., Vijaykrishna, D., Poon, L.L.M., Peiris, J.S.M., Smith, G.J.D., Chen, H., Guan, Y., 2006. Prevalance and genetic diversity of coronaviruses in bats from China. J. Virol. 80 (15), 7481–7490.

van den Born, E., Gultyaev, A.P., Snijder, E.J., 2004. Secondary structure and function of the 5′-proximal region of the equine arteritis virus RNA genome. RNA 10 (3), 424–437.

van den Born, E., Posthuma, C.C., Gultyaev, A.P., Snijder, E.J., 2005. Discontinuous subgenomic RNA synthesis in arteriviruses is guided by an RNA hairpin structure located in the genomic leader region. J. Virol. 79 (10), 6312–6324.

van der Hoek, L., Pyrc, K., Jebbink, M.F., Vermeulen-Oost, W., Berkhout, R.J., Wolthers, K.C., Wertheim-van Dillen, P.M., Kaandorp, J., Spaargaren, J., Berkhout, B., 2004. Identification of a new human coronavirus. Nat. Med. 10 (4), 368–373.

van der Most, R.G., Bredenbeek, P.J., Spaan, W.J., 1991. A domain at the 3′ end of the polymerase gene is essential for encapsidation of coronavirus defective interfering RNAs. J. Virol. 65 (6), 3219–3226.

Vijaykrishna, D., Smith, G.J., Zhang, J.X., Peiris, J.S., Chen, H., Guan, Y., 2007. Evolutionary insights into the ecology of coronaviruses. J. Virol. 81 (8), 4012–4020.

Wang, Y., Zhang, X., 2000. The leader RNA of coronavirus mouse hepatitis virus contains an enhancer-like element for subgenomic mRNA transcription. J. Virol. 74 (22), 10571–10580.

Woo, K., Joo, M., Narayanan, K., Kim, K.H., Makino, S., 1997. Murine coronavirus packaging signal confers packaging to nonviral RNA. J. Virol. 71 (1), 824–827.

Woo, P.C.Y., Lau, S.K.Y., Chu, C.M., 2005. Characterization and complete genome sequence of a novel coronavirus, coronavirus HKU1, from patients with pneumonia. J. Virol. 79 (2), 884–895.

Woo, P.C.Y., Lau, S.K.P., Li, K.S., Poon, R.W., Wong, B.H., Tsoi, H.W., Yip, B.C., Huang, Y., Chan, K.H., Yuen, K.Y., 2006. Molecular diversity of coronaviruses in bats. Virology 351 (1), 180–187.

Woo, P.C.Y., Wang, M., Lau, S.K.P., Xu, H., Poon, R.W., Guo, R., Wong, B.H., Gao, K., Tsoi, H.W., Huang, Y., Li, K.S., Lam, C.S., Chan, K.H., Zheng, B.J., Yuen, K.Y., 2007. Comparative analysis of twelve genomes of three novel group 2c and group 2d coronaviruses reveals unique group and subgroup features. J. Virol. 81 (4), 1574–1585.

Woo, P.C.Y., Lau, S.K.P., Lam, C.S.F., Lai, K.K.Y., Huang, Y., Lee, P., Luk, G.S.M., Dyrting, K.C., Chan, K.H., Yuen, K.Y., 2009. Comparative analysis of complete genome sequences of three avian coronaviruses reveals a novel group 3c coronavirus. J. Virol. 83 (2), 908–917.

Wu, H.Y., Ozdarendeli, A., Brian, D.A., 2006. Bovine coronavirus 5′-proximal genomic acceptor hotspot for discontinuous transcription is 65 nucleotides wide. J. Virol. 80 (5), 2183–2193.

Zhang, J., Guy, J.S., Snijder, E.J., Denniston, D.A., Timoney, P.J., Balasuriya, U.B., 2007. Genomic characterization of equine coronavirus. Virology 369 (1), 92–104.

Zuker, M., 2003. Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Res. 31 (13), 3406–3415.

Züst, R., Miller, T.B., Goebel, S.J., Thiel, V., Masters, P.S., 2008. Genetic interactions between an essential 3′ *cis*-acting RNA pseudoknot, replicase gene products, and the extreme 3′ end of the mouse coronavirus genome. J. Virol. 82 (3), 1214–1228.