Title Page:

Article Type: Bioinformatics

# Title: Identification of consensus hairpin loop structure among the negative sense sub-genomic RNAs of SARS-CoV-2

**Naveen Prakash Bokolia**[1], ***Ravisekhar Gadepalli**[1*]

[1]Viral Research and Diagnostic Laboratory, Microbiology Department, All India Institute of Medical Sciences, Jodhpur (342001), India

*Corresponding Author: Phone: +91-9680009207; Email: gadepallirs@gmail.com

**Keywords:** SARS-CoV-2, subgenomic RNA, Secondary structure, Transcriptional pausing, Hairpin loop structure

**Abstract:**

SARS-CoV-2 is the causative agent of worldwide pandemic disease COVID-19. SARS-CoV-2 bears positive sense RNA genome, that have organized and complex pattern of

replication/transcription process including the generation of subgenomic RNAs. Transcription regulatory sequences (TRS) have important role in the pausing of replication/transcription and generation of subgenomic RNAs. In the present bioinformatics analysis a consensus secondary structure was identified among negative sense subgenomic RNAs at the adjacent of initiation codon. This study proposed that consensus structured domain could be involve in mediating the long pausing of replication/transcription complex and responsible for subgenomic RNA production.

**Introduction:**

COVID-19 is the worldwide pandemic disease caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) that carries a positive sense, single strand RNA genome with the size of 29.9 kb (Kim et al., 2020; Zhu et al., 2020). The RNA genome undergoes complex pattern of replication/transcription process inside the host environment. The SARS-CoV-2 enters the host cell and undergoes replication of positive sense RNA genome, where negative sense RNA genome is generated, that subsequently serve as template for positive sense RNA in order to translate the viral proteins and packaging of RNA genome in to virion particles. SARS-CoV-2 RNA genome bears conserved transcriptional regulatory sequences (TRS) of 6-7 nucleotides(Yang et al., 2020). The TRS is present at the immediate upstream of initiation codon, where replication/transcription complex is paused (Alexandersen et al., 2020; Kim et al., 2020). RdRp has the important property to backtracking that causes long pausing at specific site (Dulin et al., 2015; Malone et al., 2021). Therefore, in addition to complete RNA genome, a nested set of negative sense subgenomic RNAs is also generated including S subgenomic, 3a subgenomic, E subgenomic, M subgenomic, 6 subgenomic, 7a subgenomic, 7b subgenomic, 8 subgenomic, N subgenomic and 10 subgenomic. These canonical subgenomic RNAs are thought to be generated through the complex mechanism that involves pausing of negative sense RNA synthesis by RNA dependent

48 RNA polymerase (RdRp) (Alexandersen et al., 2020; Kim et al., 2020; Mohammadi-
49 Dehcheshmeh et al., 2021; Yang et al., 2020).

50 Although, the conserved role of TRS has been identified/proposed in the pausing of
51 transcription, however, the role of RNA secondary structure has not been identified in
52 this perspective. Therefore, in this study investigation of conserved secondary structure
53 pattern was performed nearby the initiation codon that might be important in pausing
54 of transcription during negative sense RNA synthesis of the SARS-CoV-2 genome.

55 The analysis of negative sense subgenomic RNAs was performed in this study because of
56 following important reason. According to previous research model, it was suggested that
57 strand exchange or 5′ leader (5′ UTR) sequence to body (RNA) fusion occurs at TRS
58 during negative sense RNA synthesis (Kim et al., 2020). Therefore, the consensus
59 secondary structural motifs nearby the TRS might be important in several perspectives.
60 In this study a bioinformatics analysis of negative sense subgenomic RNAs was
61 performed that revealed a consensus hairpin loop secondary structure. This study
62 proposes that consensus hairpin loop secondary structure during the synthesis of
63 negative sense RNA mediates important role in transcriptional pausing and generation
64 of subgenomic RNAs.

65 **Results:**

66 **The negative sense sub-genomic RNAs bear conserved hairpin domain at immediate**
67 **downstream of initiation codon**

68 Study by Alexandersen et al., 2021, determined the presence and abundance of
69 subgenomic RNAs from the COVID-19 patient samples. Authors mapped the NGS reads
70 to subgenomic RNAs in order to identify the abundance of particular subgenomic RNA.
71 NGS reads mapped results revealed that Orf7a and N subgenomic were in high
72 abundance whereas Orf8, Orf6, and E subgenomic were relatively low (in an increasing

order). These subgenomic RNAs are present several fold higher in comparison to whole genome fragments. Therefore, the present study seek to identify the possible conserved regions nearby the transcription regulatory sequences (TRS) or initiation codon those could mediate significant roles in the generation or expression of sub-genomic RNAs.

The negative sense RNA is firstly generated through the replication/transcription of positive sense RNA genome. The negative sense RNA will serve as template for positive sense RNA genome and subgenomic messenger RNAs (or subgenomic RNAs) (Sawicki et al., 2007; Yang et al., 2020).

Therefore, in order to identify the consensus secondary structure, negative sense subgenomic RNAs were aligned by using LocARNA webserver. The significant length of upstream and downstream region (with respect to initiation codon) was considered for the analysis, with the total length of ~420 nt. Sequence alignment revealed a consensus hairpin domain (~35 to 60 nt; depending on the particular subgenomic RNA) among different negative sense sub-genomic RNAs (Figure 1, Table 1). This domain is present at the immediate downstream of initiation codon. The important features could be noted as the distance between initiation codon and consensus hairpin domain (Table 1). The structured hairpin domain of negative sense subgenomic RNAs might be of intrinsic functioning in either pausing /backtracking of RdRp at the immediate downstream of initiation codon. The conserved hairpin domain could also have role in template switching during transcription, although the role of conserved TRS has only been identified in recombination events (Yang et al., 2021). The conserved TRS is AACGAAC, that is highlighted in green (as GUUCGUU; negative sense) in the analyzed sequences of negative sense subgenomic RNAs (supplementary data file).

**The Orf7a subgenomic, N subgenomic, E subgenomic, Orf6 subgenomic and Orf8 subgenomic has distinct secondary structure domain:**

98 Since the sequence alignment of subgenomic RNAs revealed single hairpin domain.
99 However, it was required to perform secondary structure prediction of subgenomic RNA
100 sequence to examine the possibility of consensus secondary structure formation. The
101 secondary structure prediction was performed by using Vienna RNA webserver, and
102 determined whether particular subgenomic RNA adopts a specific hairpin structure or
103 not. Data obtained from sequence alignment and secondary structure prediction
104 indicates that Orf7a subgenomic, N subgenomic, E subgenomic, Orf6 subgenomic and
105 Orf8 subgenomic bears distinct and long hairpin domain with high probability of
106 formation (Table 1, Figure 2A). These structural features could possibly be correlated
107 with the previous study by Alexandersen et al; which described the high abundance of
108 Orf7a subgenomic and N subgenomic RNAs whereas Orf8 subgenomic, Orf6
109 subgenomic, and E subgenomic were relatively low (Alexandersen et al., 2020). In
110 addition, the remaining subgenomic RNAs (S subgenomic, Orf7b, Orf3a and M
111 subgenomic) were reported to be very low or near zero level abundance (Alexandersen et
112 al., 2020). The underlying reason of low level of NGS reads could be correlated with
113 their respective hairpin structure (Figure 2B) and other properties mentioned in Table
114 1. These features include: hairpin loop splits in to subdomains, hairpin loop merged with
115 initiation codon, length and probability of hairpin domain formation. Overall, the
116 formation of distinct hairpin structure might be the underlying feature in the generation
117 of negative sense sub-genomic RNAs.

118 Importantly, the highest similarity (and intermediate alignment) was observed between
119 Orf7a and N subgenomic RNAs, as both subgenomic RNAs were placed together in guide
120 tree. It provides significant relevance and correlation from the previously reported high
121 abundant reads of N and Orf7a subgenomic (Alexandersen et al., 2020). In this
122 perspective, we propose that conserved secondary structural elements might be a

123 determining factor in transcriptional pausing and backtracking of RdRp that leads to
124 subgenomic RNA production.

125 **Discussion**

126 The present finding revealed the conserved hairpin structure that is present in negative
127 sense sub-genomic RNAs of SARS-CoV-2. The significance of conserved hairpin
128 structure could be of two fold reasons. Firstly, the particular hairpin loop secondary
129 structure could have intrinsic functioning in the pausing of replication/transcription of
130 negative sense subgenomic RNAs mediated by RdRp. In addition, the conserved
131 secondary structure could mediate template switching at TRS during transcription,
132 involve the joining of 5′UTR to RNA body sequence. Further experimental studies could
133 be performed to identify the much precise role of the conserved secondary structure in
134 the generation of negative sense subgenomic RNAs and maintenance of viral infection.

135 **Methods:**

136 **Extraction of SARS-CoV-2 RNA genome and subgenomic RNAs**

137 The complete genome sequence of SARS-CoV-2 isolate Wuhan-Hu-1 (NCBI Reference
138 Sequence: NC_045512.2) was used in this study, and specific regions (upstream and
139 downstream) corresponding to particular subgenomic RNA were extracted. The
140 nucleotides sequences were converted in to reverse complement that served as negative
141 sense subgenomic RNA. The negative sense RNAs were used in alignment and
142 secondary structure prediction analysis.

143 **RNA secondary structural similarity by using LocARNA webserver**

144 LocARNA-Alignment and Folding webserver align the input RNA sequences and
145 simultaneously fold them (Raden et al., 2018; Will et al., 2012). This webserver was used
146 in the present study because it folds the RNA by using very realistic energy models as

147 used by RNAfold of the Vienna RNA package. The sequences of negative sense
148 subgenomic RNAs were aligned by using LocARNA webserver. The alignment type was
149 global and in standard mode. The alignment was used to identify the consensus
150 secondary structures elements between negative sense subgenomic RNA sequences of
151 SARS-CoV-2.

152 **Secondary structure prediction by using Vienna RNA webserver**

153 Secondary structure prediction of subgenomic RNA sequences was performed by using
154 Vienna RNA webserver (Hofacker, 2003). The subset of aligned sequence region (~285
155 nt) respective to each subgenomic RNA was submitted for structure prediction by using
156 Vienna RNA webserver. The prediction results are included in supplementary data file.

157

158

159

160

161 **References:**

162 Alexandersen, S., Chamings, A., and Bhatta, T.R. (2020). SARS-CoV-2 genomic and subgenomic RNAs
163 in diagnostic samples are not an indicator of active replication. Nature Communications *11*, 6059.
164 Dulin, D., Vilfan, I.D., Berghuis, B.A., Poranen, M.M., Depken, M., and Dekker, N.H. (2015).
165 Backtracking behavior in viral RNA-dependent RNA polymerase provides the basis for a second
166 initiation site. Nucleic acids research *43*, 10421-10429.
167 Hofacker, I.L. (2003). Vienna RNA secondary structure server. Nucleic acids research *31*, 3429-3431.
168 Kim, D., Lee, J.Y., Yang, J.S., Kim, J.W., Kim, V.N., and Chang, H. (2020). The Architecture of SARS-CoV-
169 2 Transcriptome. Cell *181*, 914-921 e910.
170 Malone, B., Chen, J., Wang, Q., Llewellyn, E., Choi, Y.J., Olinares, P.D.B., Cao, X., Hernandez, C., Eng,
171 E.T., Chait, B.T., *et al.* (2021). Structural basis for backtracking by the SARS-CoV-2
172 replication–transcription complex. Proceedings of the National Academy of Sciences *118*,
173 e2102516118.
174 Mohammadi-Dehcheshmeh, M., Moghbeli, S.M., Rahimirad, S., Alanazi, I.O., Shehri, Z.S.A., and
175 Ebrahimie, E. (2021). A Transcription Regulatory Sequence in the 5' Untranslated Region of SARS-
176 CoV-2 Is Vital for Virus Replication with an Altered Evolutionary Pattern against Human Inhibitory
177 MicroRNAs. Cells *10*.

178 Raden, M., Ali, S.M., Alkhnbashi, O.S., Busch, A., Costa, F., Davis, J.A., Eggenhofer, F., Gelhausen, R.,
179 Georg, J., Heyne, S., *et al.* (2018). Freiburg RNA tools: a central online resource for RNA-focused
180 research and teaching. Nucleic acids research *46*, W25-W29.
181 Sawicki, S.G., Sawicki, D.L., and Siddell, S.G. (2007). A contemporary view of coronavirus
182 transcription. J Virol *81*, 20-29.
183 Will, S., Joshi, T., Hofacker, I.L., Stadler, P.F., and Backofen, R. (2012). LocARNA-P: accurate boundary
184 prediction and improved detection of structural RNAs. RNA (New York, N.Y.) *18*, 900-914.
185 Yang, Y., Yan, W., Hall, A.B., and Jiang, X. (2020). Characterizing Transcriptional Regulatory
186 Sequences in Coronaviruses and Their Role in Recombination. Molecular Biology and Evolution *38*,
187 1241-1248.
188 Yang, Y., Yan, W., Hall, A.B., and Jiang, X. (2021). Characterizing Transcriptional Regulatory
189 Sequences in Coronaviruses and Their Role in Recombination. Mol Biol Evol *38*, 1241-1248.
190 Zhu, N., Zhang, D., Wang, W., Li, X., Yang, B., Song, J., Zhao, X., Huang, B., Shi, W., Lu, R., *et al.* (2020).
191 A Novel Coronavirus from Patients with Pneumonia in China, 2019. The New England journal of
192 medicine *382*, 727-733.

193

194

| Sr. No. | Type of subgenomic RNA | Location of hairpin structure (from the downstream of initiation codon) | Length of hairpin loop structure with probability | Hairpin structure is distinct or not (merged with different loop carrying initiation codon) | Sequence alignment matches with predicted structure (probability) | Probability of formation of distinct secondary structure |
|---|---|---|---|---|---|---|
| 1 | S subgenomic | Includes ~28 nt of coding region | ~55 nt | No (merged with initiation codon) | Yes | Less |
| 2 | *3a subgenomic | ~10 nt | ~58 nt | Yes | Yes | Moderate |
| 3 | ***E subgenomic | ~47 nt | ~52 nt | Yes | Yes | High |
| 4 | *M subgenomic | ~8 nt | Splits in three small hairpin loops | No (first loop merged with initiation codon) | No (three distinct hairpin structures) | Moderate |
| 5 | **Orf6 subgenomic | ~41 nt | ~47 nt | Yes | Yes | High |
| 6 | ****Orf7a subgenomic | ~30 nt | ~35 nt | Yes | Yes | High |
| 7 | Orf7b subgenomic | Includes ~6 nt of coding sequence | Splits in two hairpin loop structure | No (first loop merged with initiation codon) | No (two distinct hairpin structures) | Moderate |
| 8 | **Orf8 subgenomic | ~20 nt | ~57 nt | Yes | Yes | Moderate |

| | | | | | | |
|---|---|---|---|---|---|---|
| 9 | ****N subgenomic | ~1st nucleotide of initiation codon | ~60 nt | Yes | Yes | Moderate |

**Table 1**: The table presents the description of analyzed negative sense subgenomic RNAs. The description is drawn from the combined output of sequence alignment followed by secondary structure prediction results. The number of asterisks (*) denotes to the abundance level for each subgenomic RNA detected by NGS reads mapped to the sequences and no asterisk is given to particular subgenomic that was detected at very low or near zero level) (Alexandersen et al., 2020).

**Figure legend**:

**Figure 1**

The figure presents the output of alignment result of negative sense subgenomic RNAs that was done through the LocARNA webserver. The alignment result revealed the consensus sequence and hairpin structure (with double headed arrow) among negative sense subgenomic RNAs. This consensus structure is present at the immediate downstream of initiation codon or TRS (mentioned with double headed arrow).

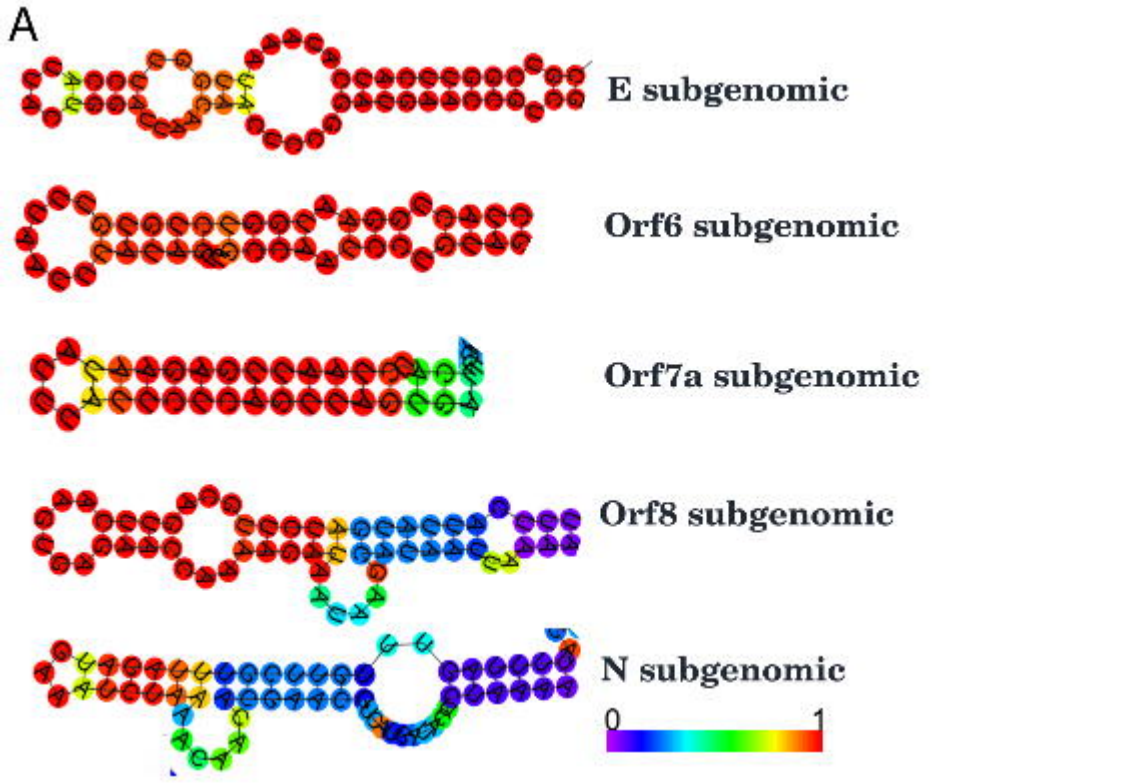
**Figure 2**

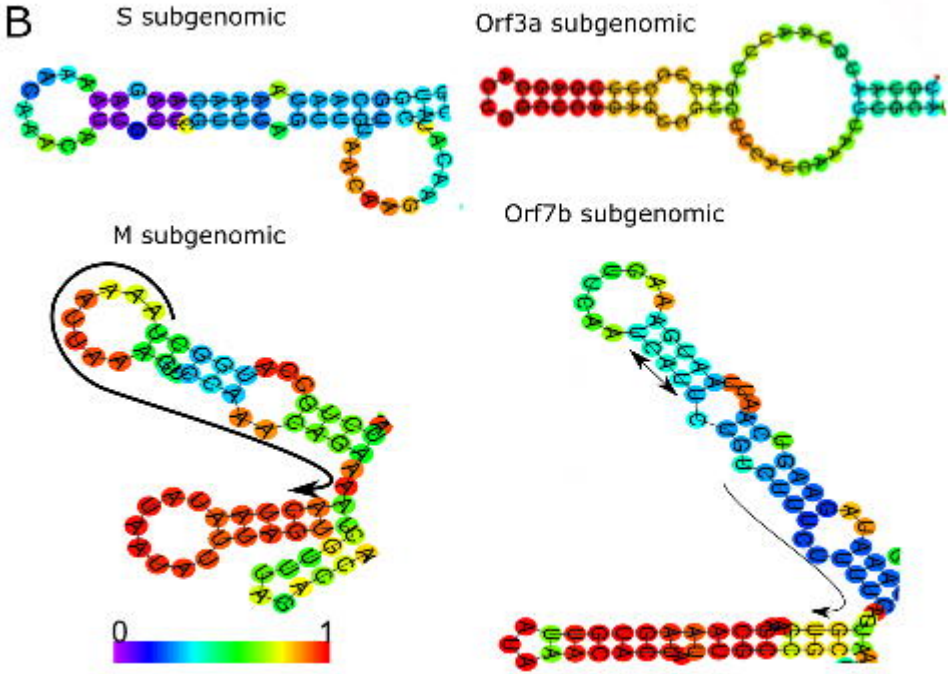
Secondary structure of consensus sequence of respective negative sense subgenomic RNAs. The secondary structure of each sequence was predicted by using Vienna RNA webserver. The respective portion of hairpin loop is shown with probability in color coding from 0 to 1. Secondary structure of consensus alignment sequence of each subgenomic RNA is shown in panel A and B. Panel A involves consensus secondary structure from subgenomic RNAs those have been detected or mapped at considerable level through NGS reads mapping (Alexandersen et al., 2020). Panel B involves consensus structure of subgenomic RNAs those have been detected at very low or zero

217 level through NGS reads mapping. The complete secondary structure of each
218 subgenomic RNA is provided in supplementary data file.

219

220


A
E subgenomic
Orf6 subgenomic
Orf7a subgenomic
Orf8 subgenomic
N subgenomic
0
1


B
S subgenomic
Orf3a subgenomic
M subgenomic
Orf7b subgenomic
0
1

```
                                   ...................(((((((.................
Orf6a   UCACAAGUAGCGAG---------------UGUU-AUCAGUGCCAAGAAAAGAAUAAUUUU 44
Orf7a   --------AGUUUA---------------GGUGAAACUGAUCUGGCACGUAACUGAUAGA 37
S       GCAAAAUAAAC-AC---------------CAUCAUUAAAUGGUAGGACAGGGUUAUCA-- 42
Orf7b   GGGUCAUCAACUACAUAUGGUUGAUGUUGAGUACAUGACUGUAAACUACAUUCU------ 54
Orf8b   AGGAAACUGUAUAAUUACCGAUAUCGAUGUACUGAAUGGGUGA----------------- 43
E       AU------------------------------CAGGAACUCUAGAAGAA------------ 19
M       AUUCUGUAAACAGCAGCAAGCACAAAACAAGCUAAAGUUA----------CUGGCCAUAA 50
Orf3a   A-------------------------------------CCCUUGGAGAGUG---------- 14
N       GAAC---------------------------------GCCUUGUCCUCGAGGGAAUUUAAGGU 30
        .........10........20........30........40........50........6

        ....................))))))).................(((((((....)))))
Orf6a   CAUGUUCGUUUAAUCAAUCUCCAUUGGUUGCUCUUCAUCUAAUUGAG-AAUAUUUAUU-C 102
Orf7a   CGUGUUUUAC--------GCCGUCAGGACAAGCAAAAGCAAAUUGAG-UGCUAAAGCAAG 88
S       ---AACCUCUUAGUA-----CCAUUGGUCCCAGAGACAUGUAU--AG-CAUGGAACCAAG 91
Orf7b   UGGUGAAAUGCAG-----CUACAGUUGUGAUGAUUCCUAAGAAAACA-AGAAAUUUC-AU 107
Orf8b   ------UUUAGAACCAGCCUCAUCCACGCACAAUUCAAUUAAAGGUGCUGAUUUUCUAGC 97
E       ------------------UUCAGAUUUUUAACACGAGAGUAAACGUAAAAAGAAGGUUUU 61
M       CAGCCAGAGGAAAAU-------UAACUUAAUUAUAUACAAAAACCUAUUCCUGUUGGCAU 103
Orf3a   -CUAGUUGCCAUCUCUUUUUGAGGGUUAU-------------------------------- 42
N       CUUCCUUGCCAUGUUGAG-UGAGAGCGGU-------------------------------- 58
        0........70........80........90........100.......110.......1

        )))...................................(((((((((...............
Orf6a   UCAG-UUAGUGACUUAGAU---AAAUUUUUAAUUAUGAGGUUUAU-------GAUGUAAU 151
Orf7a   UCAG---UGCAAAUUUGUUAUCAGCUAGAGGAUGAAAUGGUGAAUUGCCCUCGUAUGUUC 145
S       UAAC-AUUGGAAA------AGAAAGGUAAGAACAAGUCCUGAGUUG--------------- 130
Orf7b   GUUC-GUUUAGGCGUGACAAGUUUC----------AUUAUGAUCUUG--------------- 143
Orf8b   UCCU-ACUCUAAUAUACCAUUUAGAAUAGAAGUGAAUAGGACACGGG------------U 144
E       ACA--AGACUCACGU------------------UAACAAUAUUUGCAG------------- 88
M       AGGCAAAUUG--------UAGAAGACAAAUCCAUGUAAGGAAUAG---------------- 140
Orf3a   ---GAUUUUGGAAGCGCUCUGAAAAACAGCAAGAAGUGCAACGCCAACAAUAAGCCAUCC 99
N       ---GAACCA----AGACGCAGUAUUAUUGGGUAAACCUUGGGGCCGAC------------ 99
        20.......130.......140.......150.......160.......170.......1

        ..........................))))))))).......................((
Orf6a   CAAGAUUCCA-AAUGGAA-ACUUUAAAAGUCCUCAUAAUAAUUAGUAAUAUCUCUGCUAU 209
Orf7a   CAGAAGAGC--AAGGUUC-UUUUAAAAGUACUGUUGUACCUCUAACA-------CACUCU 195
S       -----------------A-AUGUAAAACUGAGGAUCUGAAAACUUUGUCAGGGUAAUAAA 172
Orf7b   CAGUUC-----AAGUGAG-AACCAAAAGAUAAUAAGCAUAAUUAAAACAAGG--AAUAGC 195
Orf8b   CAUCAACUAC-AUAUGGU-UGAUGUUGAGUA---------------------------CAU 176
E       CAGUACGCAC-ACAAUCGAAGCGCAGUAAGGAUGGCUAGUGUAACUAGCAAGAAUACCAC 147
M       -GAAACCUAU-UACUAGG--UUCCAUUGUUCA-----------AGGAGCUUUUUAAGCUC 185
Orf3a   GAAAGGGAGUGAGGCUUGUAUCGGUAUCGUUGC--------------------------AG 134
N       ----------GUUGUUUUGAUCGC--GCCCCAC--------------------------UG 122
        80.......190.......200.......210.......220.......230.......2

        (((((.............................................)))))))...
Orf6a   AGUAAC--CUGAAAGU-C-AACGAGAU----GA-AACAUCUGUUGUCAC-UUACUGUAC- 258
Orf7a   UGGUA---GUGA-----U-AA-AGCUCACAAGU-AGCGAGUGUUAUCAG-UGCCAAGAA- 242
S       CACCACGUGUGAAAGAAU-UAGUGUAUGCAGG---GGGUAAUUGAGUUC-UGGUUGUAA- 226
Orf7b   AGAAAGGCUAAAAAGCACAAAUAGAAGUCAAUU-AAUGAAAGUUCAAUC-AUUCUGU--- 250
Orf8b   GACUG---------UAAACUACAUUCU-----U-GGUGAAAUGCAGCUA-CAGUUGUGAU 220
E       GAAAGCAAGAAAAAGAA--GUACGCUAUUAACU-AUUAACGUACCUGUCUCUUCCGAAAC 204
M       UUCAAC--GGUAA-----------------------------UAGUACCGUUGGA----- 209
Orf3a   UAGC----GCGAA--CAAAAUCUGA-AGGAGUAGCAUCCUUGAUUUCACCUUGCUUCAAA 187
N       CGUU----CUCCAUUCUGGUUACUGCCAG--UUGAAUCUGAGGGUCCACCAAACGUAA-- 174
        40.......250.......260.......270.......280.......290.......3

        ...........................................................(
Orf6a   -AAGCA--------------------------AAGCAAU----AUUGUCACUGC------U 282
Orf7a   -AAGAAUAAUUU-------UCAUGUUCGUUUAAUCAAUCUCCAUUGGUUGCUC------U 288
S       -GAUUAACA-------------------------CACUGACUAGAGA---CUA------G 251
Orf7b   ---------------------------------------------------CU------U 253
Orf8b   GAUUCCUAAGAA-------AACAAGAAAUUUCAUGUUCGUUUAGGCGUGACAA------G 267
E       GAAUGAGUACAUAAGUUCGUACUCAUCAGCUUGUGCUUACAAAGGCACGCUAGUAGUCGU 264
M       ------------------------------------------AUCUGCCAUGGCUAAAAU 227
Orf3a   GUUACAGUUCCAAUUGUGAAGAUUCUCAUAAACAAAUCCAUAAGUU------CGUUUAUG 241
N       -UGCGGGGUGCAUUUCGC-------UGAUUUUG---------GGGU------CCAUUAUC 211
        00.......310.......320.......330.......340.......350.......3

        (((.((((..((((......((((((.......)))))).........))))).))).....
Orf6a   AC-UGGAA-UGGU-------CUGUGUUUAAUUUAUAG----UU----GCCAAUCC----- 320
Orf7a   UCAUCUAA-UUGA-------GAAU-----AUUUAUUC----------UCAGUUAG----- 320
S       UGGCAAUA-AAACAAGAAAAACAAACA------UUGU----UC----GUUUAGUUGUUAA 296
Orf7b   UCUUUUGAGUGUGAAGCAAAGUGUU----AUAAACAGUAUUGC----CGCAACAA--UAA 303
Orf8b   UUUCAUUAUGAUCUUGC-AGUUC-----AAGUGAGAACCAAAA----GAUAAUA----AG 313
E       CGUCGGU-UCAUCAUAAAUUGGUUCCAUUACUGGAUUAACAACUCCGGAUGAACCG---- 319
M       UAAAGUUCCAAACAGAAAAACUAA-UAUAAUAUUUAGUUC-------GUUUAGACCAGAA 279
Orf3a   UGUAAUGUAAUUUGACUCCUUUGAGCACUG-GCUCAGAGU-CGU-----CUUCAUCAAAU 294
N       AGACAUUUUAGUUUGUUCGUUUAGAUG-AA-AUCUAAAAC-AACACGAACGUCAUGAUAC 268
        60.......370.......380.......390.......400.......410.......4

        ...........))))...........................................
Orf6a   -----------UGUAGCGACUGUAUGCAGCAAAACCUGAGUCA--CCUG----------- 356
Orf7a   -----------UGA----------------------CUUAGAUA--AAUU----------- 335
S       ---CAAGAACAUCACUAGA--------AAUAACAACUCUGUUG--UUUUCUCUAAUUAUA 343
Orf7b   ---GAAAAAUUGGAGAGUAAAGUUCU---UGAACUUCCUCUUG--UCUGAUGAACAGUUU 355
Orf8b   ---CAUAAUUAAAACAAGGAA------------------UAGCAGAAAGGCUAAAAA---- 349
E       -----------UCGAUUG----------------------------UGUG---------- 330
M       GAUCAGGAACUCUAGAAGAAUUCAGAUUUUUAACACGAGAGUAAACGUAAAAAGAAGGUU 339
Orf3a   UUGCA------GCAGGAUCCACAAGAACAACAGCCCUUGAGACAACUACAGCAACUGGUC 348
N       UCUAA------AAAGUCUUCAUAGAACGAACAACGCACUACAAGACUACCCAAUUUAGGU 322
        20.......430.......440.......450.......460.......470.......4

        ...........................................................
Orf6a   -------CUACACGCUGCGAAG----CUC----------------CCAAUUUGUAAUAAG 389
Orf7a   -------UUUAAU-UAUGAGGU----UUAUGAUGUAAUCAAGAUUCCAAAUGGAAACUUU 383
S       AGUCUACCUUUACUAAGAAGAG----AUAAAAUCAUAUCAU---------------UGAUU 385
Orf7b   AGGUGAAACUGAUCUGGCACGU----AACUGAU------------------------- 384
Orf8b   -----------GCACAAAUAGA----AGUCAAUU------AAUGAAAGUUCAAUCAUUCU 388
E       ----AAUUUGGACAUGUUCUUCAGGCUCAUCAACA-------------AUUUUAUUGUAG 373
M       UUACAAGACUCAC-----GUUAACAAUAUUGCAGCAGUACGCACACAAUCGAAGCGCAGU 394
Orf3a   AUACAGC----AAAGCAUAAUUGUCACCAUUACUAUGGCAAUCAAGCCA----------- 393
N       UCCUGGCAAUUAAUUGUAAAGGUAAACAGGAAACUGUAUAAUUACCGAUAUCGAUGUAC 382
        80.......490.......500.......510.......520.......530.......5

        ...........................................
Orf6a   AAAGCGUUC--GUG--AUGUAGCAACAGUGAUU------U------C 420
Orf7a   AAAAGUCCU--CAU--AAUAAUUAGUAAUAUCUCUGCUAU------A 420
S       UGACCUUCU--UUU----AAAGACAUAACAGCAGUACCCC------U 420
Orf7b   ---AGACGU--GUUUUACGCCGUCAGGACAAGCAAAAGCA------A 420
Orf8b   GUCUUUCUU--UUGAGUGUGAAGCAAAGUGUUA-------------U 420
E       AUGAAGAAGGUAACAUGUUCAACACCAGUGUCUGUACUCAAUUGAGU 420
M       AAGGAUGGCUAGUGUAA-----------------------CUAGCAAGA 420
Orf3a   ---------GCUAUAAAACC--------UAGCCAAAUGUAC---CAU 420
N       UGAAUGGGUGAUUUAGAACC---------AGCCUCAUCCACGCACAA 420
        40.......550.......560.......570.......580.....
```