RNA genome conservation and secondary structure in SARS-CoV-2 and SARS-related viruses

Ramya Rangan¹, Ivan N. Zheludev², Rhiju Das¹,²,³*

¹Biophysics Program, Stanford University, Stanford CA 94305
²Department of Biochemistry, Stanford University School of Medicine, Stanford CA 94305
³Department of Physics, Stanford University, Stanford CA 94305

*Corresponding author: rhiju@stanford.edu
Abstract

As the COVID-19 outbreak spreads, there is a growing need for a compilation of conserved RNA genome regions in the SARS-CoV-2 virus along with their structural propensities to guide development of antivirals and diagnostics. Using sequence alignments spanning a range of betacoronaviruses, we rank genomic regions by RNA sequence conservation, identifying 79 regions of length at least 15 nucleotides as exactly conserved over SARS-related complete genome sequences available near the beginning of the COVID-19 outbreak. We then confirm the conservation of the majority of these genome regions across 739 SARS-CoV-2 sequences reported to date from the current COVID-19 outbreak, and we present a curated list of 30 ‘SARS-related-conserved’ regions. We find that known RNA structured elements curated as Rfam families and in prior literature are enriched in these conserved genome regions, and we predict additional conserved, stable secondary structures across the viral genome. We provide 106 ‘SARS-CoV-2-conserved-structured’ regions as potential targets for antivirals that bind to structured RNA. We further provide detailed secondary structure models for the 5´ UTR, frame-shifting element, and 3´ UTR. Last, we predict regions of the SARS-CoV-2 viral genome have low propensity for RNA secondary structure and are conserved within SARS-CoV-2 strains. These 59 ‘SARS-CoV-2-conserved-unstructured’ genomic regions may be most easily targeted in primer-based diagnostic and oligonucleotide-based therapeutic strategies.
Introduction

Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) has caused a rapidly expanding global pandemic, with the COVID-19 outbreak responsible at this time for over 600,000 cases and 25,000 deaths. The emergence of this pandemic has revealed an urgent need for diagnostic and antiviral strategies targeting SARS-CoV-2. Like other coronaviruses, SARS-CoV-2 is a positive sense RNA virus, with a large RNA genome approaching nearly 30 kilobases in length. Its RNA genome contains protein-coding open reading frames (ORFs) for the viral replication machinery, structural proteins, and accessory proteins. The genome additionally harbors various cis-acting RNA elements, with structures in the 5′ and 3′ untranslated region (UTRs) guiding viral replication, RNA synthesis and viral packaging. Conserved RNA elements offer compelling targets for diagnostics. In addition, such RNA elements may be useful targets for antivirals, a concept supported by the recent development of antisense oligonucleotide therapeutics and small-molecule RNA-targeting drugs for a variety of targets across infectious and chronic diseases.\(^2\)-\(^4\)

Conserved structured RNA regions have already been shown to play critical functional roles in the life cycles of coronaviruses. Most coronavirus 5′ UTR’s harbor at least four stem loops, with many showing heightened sequence conservation across betacoronaviruses, and various stems demonstrating functional roles in viral replication.\(^5\) Furthermore, RNA secondary structure in the 5′ UTR exposes a critical sequence motif, the transcriptional regulatory sequence (TRS), that forms long-range RNA interactions necessary for facilitating the discontinuous transcription characteristic to coronaviruses.\(^6\) Beyond the 5′ UTR, the frame-shifting element (FSE) in the first protein-coding ORF (ORF1ab) includes a pseudoknot structure that is necessary for the production of ORF1a and ORF1b from two overlapping reading frames via programmed ribosomal -1 frame-shifting.\(^7\) In the 3′ UTR, mutually exclusive RNA structures including the 3′ UTR pseudoknot control various stages of the RNA synthesis pathway.\(^8\)

Beyond these canonical structured regions, the RNA structure of the SARS-CoV-2 genome remains mostly unexplored. Unbiased discovery of other conserved regions and/or structured regions in the virus has the potential to uncover further functional cis-acting RNA elements. Here, we analyze RNA sequence conservation across SARS-related betacoronaviruses and currently available SARS-CoV-2 sequences, and we identify structured and unstructured regions that are conserved in each sequence set; these intervals can provide starting points for a variety of diagnostic and antiviral development strategies (Fig. 1). To identify structured regions, we predict maximum expected accuracy structures around conserved regions and report the support of these single structures from predictions of each RNA’s structural ensemble. We additionally identify thermodynamically stable secondary structures across the whole genome, finding that currently known structures fall within these predictions, but also identifying various new candidate structured regions. We pinpoint unstructured genome intervals by identifying bases with low average base-pairing probabilities. Finally, we present secondary structure models for key RNA structural elements of SARS-CoV-2 annotated in the betacoronavirus family.

Results

RNA sequence conservation in SARS-related betacoronaviruses and SARS-CoV-2

To identify potential regions of conserved RNA secondary structure in the virus, we located stretches of the SARS-CoV-2 genome with high RNA sequence conservation across SARS-related betacoronavirus full genome sequences. By identifying regions with high RNA sequence conservation as a first step, we reasoned that we would be more likely to filter for functionally relevant structures that must be conserved through virus evolution and thereby discover targets that
are potentially less likely to develop resistance against therapeutics or to escape diagnosis as the virus evolves.

We carried out this analysis beginning with three different sequence alignments, each capturing a distinct range of complete genome sequences across the betacoronavirus genus, as follows:

1. The first multiple sequence alignment (SARSr-MSA-1) was computed using MUSCLE⁹ to align sequences curated by Ceraolo and Giorgi,¹⁰ filtered by including only the reference genome sequence NC_0405512.2¹¹ from the SARS-CoV-2 sequence set, removing the two most distal betacoronavirus sequences (MERS sequences JX869059.2 and KT368829.1), and leaving in all remaining betacoronavirus whole genome sequences. This alignment captures a range of bat coronavirus and SARS sequences with only 11 sequences. These sequences correspond well to the SARS-related group defined in Gorbalenya, Baker, et al.¹²

2. The second MSA (SARSr-MSA-2) was obtained from BLAST by searching for the 100 complete genome sequences closest to the SARS-CoV-2 reference genome. This alignment captures a larger set of SARS-CoV-2, SARS, and bat coronavirus sequences than SARSr-MSA-1 but includes many sequences with high pairwise similarity.

3. The final MSA (SARSr-MSA-3) was obtained by locating all complete genome betacoronavirus sequences from the NCBI database, removing mutually similar sequences with a 99% cutoff using CD-HIT-EST¹³, and computing an MSA with Clustal Omega¹⁴. With 180 sequences with at most 99% pairwise similarity, this MSA captures a broader set of betacoronaviruses than SARSr-MSA-1 and SARSr-MSA-2 but is more challenging to align due to higher sequence diversity.

We computed conserved regions as contiguous stretches of 15 nucleotides or longer that were 100% conserved (cutoff for SARSr-MSA-1), 98% conserved (cutoff for SARSr-MSA-2), or 54% conserved (cutoff for SARSr-MSA-3). Searching for conserved regions of 15 nucleotides or more enables the design of antisense oligonucleotides that fall within these stretches. The sequence conservation cutoffs chosen ensured that at least 75 candidate conserved stretches were used for further structure analysis for each MSA. When calculating sequence conservation at the 5´ and 3´ sequence ends of the sequence, we did not include sequences that included only leading or trailing sequence deletions up to that point to avoid sequencing artefacts.

In Fig. 2, we depict conserved regions (100% conservation cutoff, SARSr-MSA-1) alongside the genome coordinates for the reference SARS-CoV-2 sequence. We observe intervals of conservation in the 5´ UTR and 3´ UTR genome regions, as expected based on prior work demonstrating sequence conservation surrounding structured RNA elements in these regions,¹⁵ but we also noted stretches of RNA sequence conservation within some viral ORFs.

Interestingly, in SARSr-MSA-1 and SARSr-MSA-2 we found that conserved stretches overlapped with previously curated Rfam families for Coronaviridae RNA secondary structures: the frameshifting stimulation element (Rfam family RF00507), the 3´ UTR pseudoknot (Rfam family RF00165), and the 3´ stem-loop II-like motif (Rfam family RF00164) (Fig. 2). We also found overlap between conserved stretches and 5´ UTR structures for previous coronaviruses, including the original SARS virus, described in the literature, including stem loops 2-3 (SL2-3) and stem loop 5 (SL5).¹⁷ These five known RNA structures overlap with conserved regions more than expected; in 10,000 random trials, the chance that five randomly chosen intervals of these lengths all overlap with the conserved regions from SARSr-MSA-1 or SARSr-MSA-2 is less than 0.0003. The enrichment of known RNA structures in these conserved regions suggests that other conserved regions may also harbor RNA structures.

We analyzed whether sequence regions conserved across SARS and bat coronaviruses remain conserved in the current COVID-19 outbreak. We determined the conservation of conserved genome regions from SARSr-MSA-1 across SARS-CoV-2 sequences as of deposition date 03-18-
20. For this analysis, we obtained two whole-genome multiple sequence alignments, keeping only full-length genome sequences of at least 29,000 nucleotides in both cases: the first includes 103 NCBI sequences aligned by MUSCLE\textsuperscript{19} (SARS-CoV-2-MSA-1), and the second includes 739 sequences deposited to GISAID\textsuperscript{18}, aligned using MAFFT\textsuperscript{19} with SARS-CoV-2-MSA-1 as a seed (SARS-CoV-2-MSA-2). We noted conserved regions in the betacoronavirus alignment SARSr-MSA-1 were more likely to be at least 99% conserved in both SARS-CoV-2-MSA-1 and SARS-CoV-2-MSA-2 than random intervals of the same size (binomial test p-value < 1e-5). Table 1 lists these regions, which we term the SARS-related-conserved regions. These genome regions are conserved across the betacoronavirus sequences in SARSr-MSA-1 and have at least 99% sequence conservation across currently available whole-genome sequences from the SARS-CoV-2 outbreak (SARS-CoV-2-MSA-2).

Conservation percentages for SARSr-MSA-1, SARSr-MSA-2, SARS-CoV-2-MSA-1, and SARS-CoV-2-MSA-2 are included in Supplementary File 1. We expect that some diagnostic and therapeutic strategies will benefit from focusing on conserved regions across a broad range of betacoronaviruses, whereas others may benefit from focusing on regions conserved only in SARS-CoV-2; we revisit the latter category of SARS-CoV-2-unique regions below.

Predictions for structured regions in SARS-CoV-2

The intrinsic RNA structure of a conserved genome region is of interest in current medical research. On one hand, if an RNA region is sufficiently unstructured to allow binding by hybridization probes, antisense oligonucleotides may be used to disrupt these functional structures. On the other hand, stable secondary structure domains are candidates for harboring stereotyped 3D RNA folds that present targets for small-molecule drug therapeutics. Beyond identifying candidates for structured genome regions, it is additionally useful to pinpoint significantly unstructured stretches that are more likely accessible to various diagnostic and antiviral interventions including standard RT-PCR assays.

We used two approaches to make predictions for conserved structured regions in SARS-CoV-2. First, we predicted RNA structures centered on the most sequence-conserved regions of SARS-related betacoronavirus genomes (alignment SARSr-MSA-1). For each conserved stretch (at least 15 nucleotides long, 100% sequence conservation) along with 20 nucleotide flanking windows, we predicted maximum expected accuracy (MEA) secondary structures using Contrafold 2.0.\textsuperscript{20} We then sought to rank sequences based on the predicted probability that the RNA folds into the MEA structure and not other structures. For this ranking, we used the expected Matthews correlation coefficient (MCC) from each construct’s base-pairing probability matrix, computed using Arnie (https://github.com/DasLab/arnie).\textsuperscript{21} An MEA structure with a higher estimated MCC is expected to have unpaired and paired bases that better align with the construct’s predicted ensemble base-pairing probabilities, lending support to the single-structure MEA prediction. In Fig. 2 regions A-E, we display the five conserved regions with the top maximum expected accuracy (MEA) secondary structures as ranked by the expected MCC (all regions listed in Supplementary File 1).

We also sought independent methods to identify thermodynamically stable and conserved RNA structures, without initially guiding the search to focus on extremely sequence-conserved genome regions. We made predictions for structured regions using RNAz\textsuperscript{22}, beginning with the betacoronavirus alignment SARSr-MSA-1. RNAz predicts structured regions that are more thermodynamically stable than expected by comparison to random sequences of the same length and sequence composition (z-score), and additionally assesses regions by the support of compensatory and consistent mutations in the sequence alignment (SCI score). These two criteria are combined into a single P-score, which when tested empirically on a set of ncrRNAs produced a false-positive rate of 4% at a P>0.5 cutoff and 1% at a P>0.9 cutoff. To predict structured regions across the full viral genome, we scanned the SARSr-MSA-1 alignment in windows of length 120
nucleotides sliding by 40 nucleotides, predicted all RNAz hits in the plus strand at a P>0.5 cutoff, clustered the resulting hits to generate maximally contiguous loci of the genome with predicted structure, and filtered results to only include loci with at least one window with a P>0.9 structure prediction.

The RNAz approach led to the prediction of 44 structured genome loci comprising 118 windows with predicted structure (P>0.9), with these loci covering 46% of the SARS-CoV-2 genome (Figure 3). We found that five canonical RNA structures (the frameshifting element, the 3’ UTR pseudoknot, the 3’ UTR hypervariable region, 5’ UTR SL2-3, and 5’ UTR SL5) were present in these genome loci. Additionally, conserved SARS-CoV-2 regions overlap significantly with predicted RNAz loci, with 62 of 78 SARS-CoV-2 conserved intervals at a 97% sequence cutoff overlapping by at least 15 nucleotides with RNAz loci (Fig. 5). This enrichment is statistically significant (p-value<0.001 from comparisons to 10,000 random placements of conserved intervals). This enrichment also holds when considering overlaps with the conserved regions from SARSr-MSA-1; 124 of the 229 SARSr-MSA-1 conserved intervals at a 90% conservation cutoff overlap by at least 15 nucleotides with RNAz loci (p-value 0.0038). This analysis potentially expands the set of conserved structural regions of SARS-CoV-2 beyond known Rfam families and those noted in the literature (full set of RNAz loci in Supplementary File 1). Top-scoring structured windows from RNAz that overlap with conserved sequence regions in SARS-CoV-2-MSA-2 for at least 15 nucleotides are included in Table 2; we termed these SARS-CoV-2-conserved-structured regions. Overlapping intervals between the RNAz predictions and conserved sequence regions in SARSr-MSA-1 are included in Supplementary File 1.

**Conserved unstructured regions of SARS-CoV-2**

We additionally located conserved regions of the viral genome predicted to lack structure, as such regions may be desired targets for some diagnostic and therapeutic approaches. We scanned the SARS-CoV-2 reference genome in windows of length 120 nucleotides sliding by 40 nucleotides, and for each window, we predicted the base-pair probability matrix with Contrafold 2.0, using these probabilities to assemble average single-nucleotide base pairing probabilities across the genome. In Figure 3, we display the 76 stretches of the genome of length at least 15 nucleotides where every base has average base-pairing probability at most 0.4. It is interesting to note that some structured 120 nucleotide windows reported by RNAz include these unpaired stretches. We find that 58 of these unpaired stretches have at least 15 nucleotides of overlap with sequence regions that are at least 97% conserved in SARS-CoV-2-MSA-2 (Fig. 5). These unpaired stretches termed SARS-CoV-2-conserved-unstructured regions are listed in Table 3 (overlaps with SARSr-MSA-1 are included in Supplementary File 1.)

**Secondary structure models for canonical structured regions of SARS-CoV-2**

Currently known RNA structures that recur across betacoronaviruses provide potential starting points for therapeutic development targeting the SARS-CoV-2 RNA genome. Here, we include secondary structures for the 5’ UTR (Figure 4a), frame-shifting element (Figure 4b), and 3’ UTR (Figure 4c) for SARS-CoV-2, built by analyzing homology to literature-annotated structures in related betacoronaviruses. We additionally include computer-readable secondary structures in Table 4 and Supplementary File 1. A brief review of salient secondary structure features in these regions and their putative functional roles in the betacoronavirus life cycle follows.

The 5’ UTR includes five confident stem-loop structures (SL1-SL5), with structures verified by chemical mapping experiments in related coronaviruses.\(^{17,23}\) SL1 and SL2 are conserved across betacoronaviruses, with SL2 having the highest sequence conservation across the 5’ UTR.\(^{15}\) The high A-U base-pairing content in the SARS-CoV-2 SL1 sequence and the bulged nucleotides align with prior reports that SL1 is relatively thermodynamically unstable to allow for the formation of long-
range interactions.\(^{24}\) SL2 has been shown to be critical for subgenomic RNA synthesis, with mutations in its conserved pentaploa retaining the production of genome-sized RNA, but not subgenomic RNA segments.\(^{25}\) SL3, conserved only in betacoronaviruses, presents the transcription-regulating sequence (TRS) that base pairs with one of several complementary sequences in nascent negative-sense strands in a ‘copy-choice mechanism’ leading to discontinuous transcription of subgenomic mRNAs.\(^{6}\) SL4 contains a short ORF, uORF, which precedes the first longer ORF1ab of the genome. The uORF leads to attenuated transcription of ORF1ab that appears helpful but is not essential for viral replication.\(^{15}\) SL5 has been implicated in a potential role in viral packaging, and contains the AUG start codon for long ORF1ab which encodes the viral replicase/transcriptase polyprotein. The SARS-CoV-2 SL5 domain has common features with the domain in other group Iib betacoronaviruses, for instance including UUCGU pentaploons on SL5a and SL5b, and a GNRA tetraloop on SL5c.\(^{23}\) Prior DMS-probing data for Stem 5 in SARS-CoV aligned with the proposed SL5a,b,c structures.\(^{23}\) Two additional stems (SL6 and SL7) are predicted from computer modeling here, but there is only weak evidence in prior literature on other betacoronaviruses for these downstream stems being functionally important.

The frameshifting element (FSE) is located in ORF1ab and is involved in regulating a (-1) ribosomal frameshift event that is necessary for producing ORF1b. The FSE consists of a conserved pseudoknot structure that regulates the rate of ribosomal frameshifting at an upstream slippery site.\(^{7}\) This domain is nearly exactly conserved between SARS-CoV and SARS-CoV-2, suggesting a similar mechanism for ribosomal pausing and slippage between the two viruses.\(^{26}\)

The 3´ UTR contains various domains critical for regulating viral RNA synthesis and potentially translation. The most 5´ region of the 3´ UTR includes a switch-like domain involving mutually exclusive formation of a pseudoknot and stem-loop, both of which are essential for viral replication with putative roles in establishing the kinetics of RNA synthesis.\(^{6,27}\) The hyper-variable region (HVR) is not essential for viral RNA synthesis, as this can be removed while allowing for viral replication in tissue culture; however, viruses without this domain have lower pathogenicity in mice.\(^{28}\) This domain contains a completely conserved octonucleotide sequence with unconfirmed functional significance. The stem-loop II-like motif (s2m) is another subregion of the HVR that is conserved in SARS-CoV-2 and other coronaviruses. A crystal structure of the SARS s2m domain has been shown to be homologous to an rRNA loop that binds translation initiation proteins, leading this domain to have a proposed role in recruiting host translational machinery.\(^{29}\) The domain has been proposed to be a selfish element due to its recurrence in numerous virus families outside the Coronaviridae, but its function is not well understood.\(^{30}\)

**Discussion**

Understanding the RNA structure of the SARS-CoV-2 genome can guide RNA-targeting interventions and diagnostics. Here we have presented an initial analysis of RNA sequence conservation across betacoronaviruses and current SARS-CoV-2 sequences, predictions for structured and unstructured domains of the viral RNA genome, and homology-derived secondary structure models for classic structured elements of the SARS-CoV-2 genome: the 5´ UTR, the frame-shifting element, and the 3´ UTR. Our hope is that these steps will provide a useful starting point for efforts to develop antivirals and diagnostics that depend on targeting either structured or unstructured viral genomic regions.

The abundant RNA structures involved in the replication cycle of betacoronaviruses present ample opportunities for therapeutic development, but our analysis is not complete. Beyond the secondary structures highlighted here, prior work has pinpointed a variety of RNA-RNA interactions important to the betacoronavirus life cycle. Long-range interactions between the 5´ UTR and 3´ UTR have been implicated in RNA synthesis, for instance with mutations in the 5´ UTR SL1 only supporting viral
replication when co-evolving with specific mutations in the 3´ UTR.\textsuperscript{15, 31} In related coronaviruses, RNA structures that act as packaging signals have been identified in genomic ORFs.\textsuperscript{15} Such packaging signals may reside in regions identified here from our RNAz computational analysis or may have been missed.

We believe a more thorough structure/function analysis of the virus should be obtainable with recent experimental technologies that integrate multidimensional chemical mapping, electron microscopy, and computer modeling.\textsuperscript{32-33} New candidates for structured RNA elements in SARS-CoV-2 could play various functional roles, perhaps regulating viral packaging, replication, RNA synthesis, or translation initiation. To further improve these structure predictions, information about protein-binding events should be integrated, and biochemical assays can be conducted with these proteins present or even in cells. Accounting for protein-binding events will be critical to completing a picture of accessible and structured RNA sites.

The stable secondary structures predicted here present a reasonable starting point for 3D modeling of RNA-only structures in various regions, including the 5´ UTR stem loop 5, the frame-shifting element, the 3´ UTR pseudoknot, and the 3´ stem-loop II-like motif. Furthermore, these regions and novel predicted structured regions can serve as candidates for RNA-only structure determination. Such 3D structures have the potential to reveal well-defined 3D folds with conserved binding domains for small-molecule drugs, potentially presenting alternative approaches for targeting SARS-CoV-2.

**Acknowledgments**

The authors acknowledge support from the National Science Foundation Graduate Research Fellowship Program under grants no. 1650114 (R.R.), a Stanford Graduate Fellowship (I.N.Z.), and NIH grant MIRA R35 GM122579 (to R.D.). We would like to thank Dr. Rachel H. Saluti, Dr. Edward A. Pham, and Dr. Jeffrey S. Glenn for providing advice on the conserved, structured intervals desirable for antisense oligonucleotide design; Hannah Wayment-Steele for useful discussions on secondary structure modeling; and Andrew M. Watkins for updates to the RiboDraw software for rendering secondary structures.
Figures

**Figure 1:** We aim to provide a series of genome regions in SARS-CoV-2 that are useful for a variety of diagnostic and therapeutic strategies, including regions that are conserved in SARS-related betacoronaviruses and SARS-CoV-2 sequences (Table 1), regions that are structured and conserved in SARS-CoV-2 sequences (Table 2), and regions that are unstructured and conserved in SARS-CoV-2 sequences (Table 3).
Figure 2: In black we annotate SARSr-MSA-1 conserved regions of the genome, superimposed on SARS-CoV-2 genome ORFs. We depict the top secondary structures as ranked by Matthews correlation coefficient that overlap with these conserved regions, ordered from A to E. Regions A to E are annotated on the genome in yellow and are located at genome positions: A:13743-13798, B:17511-17566, C:28990-29054, D:172-236, E:26-109. Secondary structures are colored by sequence conservation in SARSr-MSA-1 (cyan = more conserved, purple = less conserved). In magenta are depicted curated Rfam families present in coronaviruses, including the frame-shifting element (FSE), the 3’ UTR pseudoknot (PK3), and the 3’ stem-loop II-like motif (s2m).
**Figure 3:** Structured intervals predicted from RNAz (cyan) and unstructured intervals (yellow) on the genome ORFs for SARS-CoV-2.\textsuperscript{34} (A-C) highlight the three secondary structures for windows that do not overlap with known Rfam or literature-annotated structures with the highest P-value scores from RNAz (all P>0.9). These windows are located at genome positions 14207-14366 (A), 17126-17245 (B), and 26176-26295 (C). Secondary structures are colored by sequence conservation (cyan = more conserved, purple = less conserved).
Figure 4. Secondary structure diagrams for A) 5’ UTR, B) Frameshift element, C) 3’ UTR. Nucleotides are black if 100% conserved in the SARS, bat, and SARS-CoV-2 sequences in SARSr-MSA-1, and grey otherwise. Special labeled domains are in boldface. Structures are based primarily on manual identification of homology with literature coronavirus structure models. Figures prepared in RiboDraw (https://github.com/ribokit/RiboDraw).
Figure 5. We depict the predicted number of structured, unstructured and conserved intervals for a choice of sequence conservation cutoffs. The SARS-related conserved intervals are all regions of at least 15 nucleotides with each position at least 90% conserved across an alignment of SARS, bat coronavirus, and SARS-CoV-2 sequences (SARSr-MSA-1). The SARS-CoV-2 intervals are regions of at least 15 nucleotides with each position at least 97% conserved across an alignment of currently available SARS-CoV-2 sequences (SARS-CoV-2-MSA-2). Structured intervals are loci predicted from RNAz with some loci containing multiple RNAz windows, and unstructured intervals are stretches of at least 15 nucleotides where all bases have base-pairing probability at most 0.4. All interval intersections are required to have at least 15 nucleotide overlaps, with the number of overlapping intervals listed for each interval type involved in the intersection. Top-scoring structured intervals conserved in SARS-CoV-2 sequences (green) are listed in Table 2. Top-scoring unstructured intervals conserved in SARS-CoV-2 sequences (blue) are listed in Table 3.
Table 1: SARS-related-conserved. Conserved regions across SARSr-MSA-1 and SARS-CoV-2-MSA-2. All intervals are at least 90% conserved across the SARS and bat coronavirus sequences in SARSr-MSA-1, have length at least 15 nucleotides, and have every position at least 99% conserved in current GISAID SARS-CoV-2 sequences (SARS-CoV-2-MSA-2). Sequence intervals are relative to the reference genome NC_045512.2.
Table 1: Secondary structure and conservation analysis of SARS-CoV-2

<table>
<thead>
<tr>
<th>Name</th>
<th>Sequence interval</th>
<th>Sequence</th>
<th>Secondary structure</th>
<th>z-score</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>SARS-CoV-2-conserved-1</td>
<td>918-1037</td>
<td>AAUGUCAUGUCAAGUGAAUAGAGGGGUUUUACAGGCGCC</td>
<td>((...((((((...))))))))...((((((...))))))))...)</td>
<td>-4.38</td>
<td>0.999</td>
</tr>
<tr>
<td>SARS-CoV-2-conserved-2</td>
<td>1420-14325</td>
<td>AAUGUCAUGUCAAGUGAAUAGAGGGGUUUUACAGGCGCC</td>
<td>((...((((((...))))))))...((((((...))))))))...)</td>
<td>-1.99</td>
<td>0.999</td>
</tr>
<tr>
<td>SARS-CoV-2-conserved-3</td>
<td>17125-17244</td>
<td>AAUGUCAUGUCAAGUGAAUAGAGGGGUUUUACAGGCGCC</td>
<td>((...((((((...))))))))...((((((...))))))))...)</td>
<td>-6.85</td>
<td>0.999</td>
</tr>
<tr>
<td>SARS-CoV-2-conserved-4</td>
<td>7011-7130</td>
<td>AAUGUCAUGUCAAGUGAAUAGAGGGGUUUUACAGGCGCC</td>
<td>((...((((((...))))))))...((((((...))))))))...)</td>
<td>-1.96</td>
<td>0.999</td>
</tr>
<tr>
<td>SARS-CoV-2-conserved-5</td>
<td>26135-26254</td>
<td>AAUGUCAUGUCAAGUGAAUAGAGGGGUUUUACAGGCGCC</td>
<td>((...((((((...))))))))...((((((...))))))))...)</td>
<td>-5.44</td>
<td>0.999</td>
</tr>
<tr>
<td>SARS-CoV-2-conserved-6</td>
<td>26175-26294</td>
<td>AAUGUCAUGUCAAGUGAAUAGAGGGGUUUUACAGGCGCC</td>
<td>((...((((((...))))))))...((((((...))))))))...)</td>
<td>-4.44</td>
<td>0.999</td>
</tr>
<tr>
<td>SARS-CoV-2-conserved-7</td>
<td>6251-6370</td>
<td>AAUGUCAUGUCAAGUGAAUAGAGGGGUUUUACAGGCGCC</td>
<td>((...((((((...))))))))...((((((...))))))))...)</td>
<td>-2.52</td>
<td>0.998</td>
</tr>
<tr>
<td>SARS-CoV-2-conserved-8</td>
<td>40-157</td>
<td>AAUGUCAUGUCAAGUGAAUAGAGGGGUUUUACAGGCGCC</td>
<td>((...((((((...))))))))...((((((...))))))))...)</td>
<td>-2.91</td>
<td>0.998</td>
</tr>
<tr>
<td>SARS-CoV-2-conserved-9</td>
<td>26491-26607</td>
<td>AAUGUCAUGUCAAGUGAAUAGAGGGGUUUUACAGGCGCC</td>
<td>((...((((((...))))))))...((((((...))))))))...)</td>
<td>-3.52</td>
<td>0.998</td>
</tr>
<tr>
<td>SARS-CoV-2-conserved-10</td>
<td>25895-26014</td>
<td>AAUGUCAUGUCAAGUGAAUAGAGGGGUUUUACAGGCGCC</td>
<td>((...((((((...))))))))...((((((...))))))))...)</td>
<td>-4.4</td>
<td>0.997</td>
</tr>
<tr>
<td>SARS-CoV-2-conserved-11</td>
<td>6211-6330</td>
<td>AAUGUCAUGUCAAGUGAAUAGAGGGGUUUUACAGGCGCC</td>
<td>((...((((((...))))))))...((((((...))))))))...)</td>
<td>-2.76</td>
<td>0.996</td>
</tr>
<tr>
<td>SARS-CoV-2-conserved-12</td>
<td>26215-26334</td>
<td>AAUGUCAUGUCAAGUGAAUAGAGGGGUUUUACAGGCGCC</td>
<td>((...((((((...))))))))...((((((...))))))))...)</td>
<td>-3.49</td>
<td>0.996</td>
</tr>
<tr>
<td>SARS-CoV-2-conserved-13</td>
<td>6211-6330</td>
<td>AAUGUCAUGUCAAGUGAAUAGAGGGGUUUUACAGGCGCC</td>
<td>((...((((((...))))))))...((((((...))))))))...)</td>
<td>-1.03</td>
<td>0.996</td>
</tr>
<tr>
<td>SARS-CoV-2-conserved-14</td>
<td>16645-16804</td>
<td>AAUGUCAUGUCAAGUGAAUAGAGGGGUUUUACAGGCGCC</td>
<td>((...((((((...))))))))...((((((...))))))))...)</td>
<td>-1.17</td>
<td>0.997</td>
</tr>
<tr>
<td>SARS-CoV-2-conserved-15</td>
<td>7051-7170</td>
<td>AAUGUCAUGUCAAGUGAAUAGAGGGGUUUUACAGGCGCC</td>
<td>((...((((((...))))))))...((((((...))))))))...)</td>
<td>0.99</td>
<td>0.996</td>
</tr>
<tr>
<td>SARS-CoV-2-conserved-16</td>
<td>158-277</td>
<td>AAUGUCAUGUCAAGUGAAUAGAGGGGUUUUACAGGCGCC</td>
<td>((...((((((...))))))))...((((((...))))))))...)</td>
<td>0.99</td>
<td>0.996</td>
</tr>
<tr>
<td>SARS-CoV-2-conserved-17</td>
<td>23977-24096</td>
<td>AAUGUCAUGUCAAGUGAAUAGAGGGGUUUUACAGGCGCC</td>
<td>((...((((((...))))))))...((((((...))))))))...)</td>
<td>2.87</td>
<td>0.996</td>
</tr>
<tr>
<td>SARS-CoV-2-conserved-18</td>
<td>25855-25974</td>
<td>AAUGUCAUGUCAAGUGAAUAGAGGGGUUUUACAGGCGCC</td>
<td>((...((((((...))))))))...((((((...))))))))...)</td>
<td>2.62</td>
<td>0.994</td>
</tr>
<tr>
<td>SARS-CoV-2-conserved-19</td>
<td>2943-3062</td>
<td>AAUGUCAUGUCAAGUGAAUAGAGGGGUUUUACAGGCGCC</td>
<td>((...((((((...))))))))...((((((...))))))))...)</td>
<td>-4.25</td>
<td>0.994</td>
</tr>
<tr>
<td>SARS-CoV-2-conserved-20</td>
<td>14726-14845</td>
<td>AAUGUCAUGUCAAGUGAAUAGAGGGGUUUUACAGGCGCC</td>
<td>((...((((((...))))))))...((((((...))))))))...)</td>
<td>-0.32</td>
<td>0.994</td>
</tr>
<tr>
<td>SARS-CoV-2-conserved-21</td>
<td>26375-26490</td>
<td>AAUGUCAUGUCAAGUGAAUAGAGGGGUUUUACAGGCGCC</td>
<td>((...((((((...))))))))...((((((...))))))))...)</td>
<td>-3.02</td>
<td>0.994</td>
</tr>
<tr>
<td>SARS-CoV-2-conserved-22</td>
<td>21385-21484</td>
<td>AAUGUCAUGUCAAGUGAAUAGAGGGGUUUUACAGGCGCC</td>
<td>((...((((((...))))))))...((((((...))))))))...)</td>
<td>-2.73</td>
<td>0.994</td>
</tr>
<tr>
<td>SARS-CoV-2-conserved-23</td>
<td>2158-2277</td>
<td>AAUGUCAUGUCAAGUGAAUAGAGGGGUUUUACAGGCGCC</td>
<td>((...((((((...))))))))...((((((...))))))))...)</td>
<td>-2.27</td>
<td>0.993</td>
</tr>
<tr>
<td>SARS-CoV-2-conserved-24</td>
<td>4174-4293</td>
<td>AAUGUCAUGUCAAGUGAAUAGAGGGGUUUUACAGGCGCC</td>
<td>((...((((((...))))))))...((((((...))))))))...)</td>
<td>-1.62</td>
<td>0.993</td>
</tr>
<tr>
<td>SARS-CoV-2-conserved-25</td>
<td>2652-26647</td>
<td>AAUGUCAUGUCAAGUGAAUAGAGGGGUUUUACAGGCGCC</td>
<td>((...((((((...))))))))...((((((...))))))))...)</td>
<td>-1.33</td>
<td>0.993</td>
</tr>
<tr>
<td>SARS-CoV-2-conserved-26</td>
<td>26735-26774</td>
<td>AAUGUCAUGUCAAGUGAAUAGAGGGGUUUUACAGGCGCC</td>
<td>((...((((((...))))))))...((((((...))))))))...)</td>
<td>-3.29</td>
<td>0.993</td>
</tr>
<tr>
<td>SARS-CoV-2-conserved-27</td>
<td>5651-5770</td>
<td>AAUGUCAUGUCAAGUGAAUAGAGGGGUUUUACAGGCGCC</td>
<td>((...((((((...))))))))...((((((...))))))))...)</td>
<td>-2.31</td>
<td>0.992</td>
</tr>
<tr>
<td>SARS-CoV-2-conserved-28</td>
<td>7131-7250</td>
<td>AAUGUCAUGUCAAGUGAAUAGAGGGGUUUUACAGGCGCC</td>
<td>((...((((((...))))))))...((((((...))))))))...)</td>
<td>-2.77</td>
<td>0.992</td>
</tr>
<tr>
<td>SARS-CoV-2-conserved-29</td>
<td>10970-11089</td>
<td>AAUGUCAUGUCAAGUGAAUAGAGGGGUUUUACAGGCGCC</td>
<td>((...((((((...))))))))...((((((...))))))))...)</td>
<td>-0.12</td>
<td>0.992</td>
</tr>
</tbody>
</table>
The copyright holder for this preprint (which was not peer-reviewed) is the...
SARS-CoV-2-structured-60 7771-7890
CUJUACUGUAAUAGCGCCAGCAAGAAGAUJUAAGAAGAGCA
UAUAACCAUAGGUUAUCCUUAGAAUAUGCAUAGAAG
UAUAACCAUAGGUUAUCCUUAGAAUAUGCAUAGAAG

SARS-CoV-2-structured-61 19795-19884
UACCGGCUGAGUAGGGAUGCAUGGAGGGAUGCAUGG
AGAAGAAACAGAAGAAACAGAAGAAACAGAAGAAACAG
AAGAAACAGAAGAAACAGAAGAAACAGAAGAAACAG
AAGAAACAGAAGAAACAGAAGAAACAGAAGAAACAG

SARS-CoV-2-structured-62 6971-7090
UCUGACCAGUGGCAUAGUGGCAUAGUGGCAUAGUGG
CAUAGUGGCAUAGUGGCAUAGUGGCAUAGUGG
CAUAGUGGCAUAGUGGCAUAGUGGCAUAGUGG

SARS-CoV-2-structured-63 28998-29117
UCUGACCAGUGGCAUAGUGGCAUAGUGGCAUAGUGG
CAUAGUGGCAUAGUGGCAUAGUGGCAUAGUGG
CAUAGUGGCAUAGUGGCAUAGUGGCAUAGUGG

SARS-CoV-2-structured-64 5011-5130
UCUGACCAGUGGCAUAGUGGCAUAGUGGCAUAGUGG
CAUAGUGGCAUAGUGGCAUAGUGGCAUAGUGG
CAUAGUGGCAUAGUGGCAUAGUGGCAUAGUGG

SARS-CoV-2-structured-65 4094-4213
UCUGACCAGUGGCAUAGUGGCAUAGUGGCAUAGUGG
CAUAGUGGCAUAGUGGCAUAGUGGCAUAGUGG
CAUAGUGGCAUAGUGGCAUAGUGGCAUAGUGG

SARS-CoV-2-structured-66 1598-1717
UCUGACCAGUGGCAUAGUGGCAUAGUGGCAUAGUGG
CAUAGUGGCAUAGUGGCAUAGUGGCAUAGUGG
CAUAGUGGCAUAGUGGCAUAGUGGCAUAGUGG

SARS-CoV-2-structured-67 11090-11200
UCUGACCAGUGGCAUAGUGGCAUAGUGGCAUAGUGG
CAUAGUGGCAUAGUGGCAUAGUGGCAUAGUGG
CAUAGUGGCAUAGUGGCAUAGUGGCAUAGUGG

SARS-CoV-2-structured-68 17965-18084
UCUGACCAGUGGCAUAGUGGCAUAGUGGCAUAGUGG
CAUAGUGGCAUAGUGGCAUAGUGGCAUAGUGG
CAUAGUGGCAUAGUGGCAUAGUGGCAUAGUGG

SARS-CoV-2-structured-69 25875-25934
UCUGACCAGUGGCAUAGUGGCAUAGUGGCAUAGUGG
CAUAGUGGCAUAGUGGCAUAGUGGCAUAGUGG
CAUAGUGGCAUAGUGGCAUAGUGGCAUAGUGG

SARS-CoV-2-structured-70 14446-14565
UCUGACCAGUGGCAUAGUGGCAUAGUGGCAUAGUGG
CAUAGUGGCAUAGUGGCAUAGUGGCAUAGUGG
CAUAGUGGCAUAGUGGCAUAGUGGCAUAGUGG

SARS-CoV-2-structured-71 2118-2237
UCUGACCAGUGGCAUAGUGGCAUAGUGGCAUAGUGG
CAUAGUGGCAUAGUGGCAUAGUGGCAUAGUGG
CAUAGUGGCAUAGUGGCAUAGUGGCAUAGUGG

SARS-CoV-2-structured-72 17205-17324
UCUGACCAGUGGCAUAGUGGCAUAGUGGCAUAGUGG
CAUAGUGGCAUAGUGGCAUAGUGGCAUAGUGG
CAUAGUGGCAUAGUGGCAUAGUGGCAUAGUGG

SARS-CoV-2-structured-73 7171-7290
UCUGACCAGUGGCAUAGUGGCAUAGUGGCAUAGUGG
CAUAGUGGCAUAGUGGCAUAGUGGCAUAGUGG
CAUAGUGGCAUAGUGGCAUAGUGGCAUAGUGG

SARS-CoV-2-structured-74 14766-14885
UCUGACCAGUGGCAUAGUGGCAUAGUGGCAUAGUGG

SARS-CoV-2-structured-75 11730-11849
UCUGACCAGUGGCAUAGUGGCAUAGUGGCAUAGUGG

SARS-CoV-2-structured-76 29238-29357
UCUGACCAGUGGCAUAGUGGCAUAGUGGCAUAGUGG

SARS-CoV-2-structured-77 1558-1677
UCUGACCAGUGGCAUAGUGGCAUAGUGGCAUAGUGG

SARS-CoV-2-structured-78 4241-4333
UCUGACCAGUGGCAUAGUGGCAUAGUGGCAUAGUGG

SARS-CoV-2-structured-79 8410-9529
UCUGACCAGUGGCAUAGUGGCAUAGUGGCAUAGUGG

SARS-CoV-2-structured-80 13806-13925
UCUGACCAGUGGCAUAGUGGCAUAGUGGCAUAGUGG

SARS-CoV-2-structured-81 23537-23656
UCUGACCAGUGGCAUAGUGGCAUAGUGGCAUAGUGG

SARS-CoV-2-structured-82 9210-9329
UCUGACCAGUGGCAUAGUGGCAUAGUGGCAUAGUGG

SARS-CoV-2-structured-83 6091-6210
UCUGACCAGUGGCAUAGUGGCAUAGUGGCAUAGUGG

SARS-CoV-2-structured-84 29118-29237
UCUGACCAGUGGCAUAGUGGCAUAGUGGCAUAGUGG

SARS-CoV-2-structured-85 28958-29077
UCUGACCAGUGGCAUAGUGGCAUAGUGGCAUAGUGG

SARS-CoV-2-structured-86 29278-29397
UCUGACCAGUGGCAUAGUGGCAUAGUGGCAUAGUGG

SARS-CoV-2-structured-87 14046-14165
UCUGACCAGUGGCAUAGUGGCAUAGUGGCAUAGUGG

SARS-CoV-2-structured-88 9250-9369
UCUGACCAGUGGCAUAGUGGCAUAGUGGCAUAGUGG

SARS-CoV-2-structured-89 20415-20527
UCUGACCAGUGGCAUAGUGGCAUAGUGGCAUAGUGG

17.0.934

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.27.012906. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder. It is made available under a CC-BY 4.0 International license.
Table 2: SARS-CoV-2-conserved-structured. RNAz windows as scored by the P-value (P>0.9) that overlap with conserved intervals from SARS-CoV-2-MSA-2 (97% conservation cutoff) by at least 15 nucleotides. Sequence intervals are relative to the reference genome NC_045512.2.

<table>
<thead>
<tr>
<th>SARS-CoV-2-conserved-structured</th>
<th>RNAz Window 1</th>
<th>RNAz Window 2</th>
<th>RNAz Window 3</th>
<th>RNAz Window 4</th>
<th>RNAz Window 5</th>
<th>RNAz Window 6</th>
<th>RNAz Window 7</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACUAUUUGUUAGUGAUGGUAGUAAUGCUACUUAAUGUAGUUUCGCUAGAUACAAUGACCAUGGCUAAGGCUAAC</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>GGUGGUAGAAGCAGAACAGCAACAUAGAAGCAGGCAAA</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>UGUAGUUGUUGUUAGUGUUGUUGUUGCUAGGCUAGGCAAA</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>AGUGUAGUUGUUGUUAGUGUUGUUGCUAGGCUAGG</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>AGUGUAGUUGUUGUUAGUGUUGUUGCUAGGCUAGG</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>AGUGUAGUUGUUGUUAGUGUUGUUGCUAGGCUAGG</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>AGUGUAGUUGUUGUUAGUGUUGUUGCUAGGCUAGG</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>AGUGUAGUUGUUGUUAGUGUUGUUGCUAGGCUAGG</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>AGUGUAGUUGUUGUUAGUGUUGUUGCUAGGCUAGG</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>Name</td>
<td>Sequence interval</td>
<td>Average unpaired probability</td>
<td>Minimum unpaired probability</td>
<td>Sequence</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>------</td>
<td>-------------------</td>
<td>------------------------------</td>
<td>-----------------------------</td>
<td>----------</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SARS-CoV-2-conserved-unstructured-1</td>
<td>29074-29087</td>
<td>0.891105</td>
<td>0.763759</td>
<td>AUCAAUGUAACAC</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SARS-CoV-2-conserved-unstructured-2</td>
<td>8078-8094</td>
<td>0.82497</td>
<td>0.752542</td>
<td>CCAUGGAAAACUCAA</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SARS-CoV-2-conserved-unstructured-3</td>
<td>1359-1374</td>
<td>0.836512</td>
<td>0.716868</td>
<td>UUGUUAAAUUUAAUUG</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SARS-CoV-2-conserved-unstructured-4</td>
<td>21626-21643</td>
<td>0.857235</td>
<td>0.713387</td>
<td>ACUAAUAUACCCCUCGCA</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SARS-CoV-2-conserved-unstructured-5</td>
<td>1420-1436</td>
<td>0.796631</td>
<td>0.697304</td>
<td>CGAAUACCUAUAUGAA</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SARS-CoV-2-conserved-unstructured-6</td>
<td>18471-18484</td>
<td>0.779821</td>
<td>0.695284</td>
<td>UCAAUUUAACACC</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SARS-CoV-2-conserved-unstructured-7</td>
<td>11910-11923</td>
<td>0.766973</td>
<td>0.683262</td>
<td>AAUCAUCUCAAUAA</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SARS-CoV-2-conserved-unstructured-8</td>
<td>23960-23981</td>
<td>0.787037</td>
<td>0.677563</td>
<td>UUUUAUUUUCAAAAUAUAC</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SARS-CoV-2-conserved-unstructured-9</td>
<td>13990-14003</td>
<td>0.796334</td>
<td>0.661868</td>
<td>CAAGCUUUGUAAA</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SARS-CoV-2-conserved-unstructured-10</td>
<td>10009-10035</td>
<td>0.760204</td>
<td>0.656672</td>
<td>UCUUUAACCACCCACCAAAACCUCUAU</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SARS-CoV-2-conserved-unstructured-11</td>
<td>23700-23718</td>
<td>0.822733</td>
<td>0.654787</td>
<td>CCAUCCCCACAAAUUAC</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SARS-CoV-2-conserved-unstructured-12</td>
<td>18918-18934</td>
<td>0.832132</td>
<td>0.654059</td>
<td>UAUUGAUAUACCUAUAA</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SARS-CoV-2-conserved-unstructured-13</td>
<td>27365-27402</td>
<td>0.809693</td>
<td>0.653733</td>
<td>UAAACGCAUAUAAUUA</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SARS-CoV-2-conserved-unstructured-14</td>
<td>5773-5789</td>
<td>0.808284</td>
<td>0.65353</td>
<td>UAAACACUAUACCUCUA</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SARS-CoV-2-conserved-unstructured-15</td>
<td>23910-23932</td>
<td>0.837777</td>
<td>0.652669</td>
<td>CACAAUGUACCAAAAAUACAA</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SARS-CoV-2-conserved-unstructured-16</td>
<td>17762-17785</td>
<td>0.767426</td>
<td>0.650277</td>
<td>CUGUCUUUAUACCACCUUAUAA</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SARS-CoV-2-conserved-unstructured-17</td>
<td>25569-25582</td>
<td>0.825796</td>
<td>0.648774</td>
<td>UUCCAAAUCAAA</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SARS-CoV-2-conserved-unstructured-18</td>
<td>19569-19588</td>
<td>0.75387</td>
<td>0.648103</td>
<td>UAAACAAUAAUUGAACUU</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SARS-CoV-2-conserved-unstructured-19</td>
<td>22552-22565</td>
<td>0.773461</td>
<td>0.646815</td>
<td>UAAUAUAAACACU</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SARS-CoV-2-conserved-unstructured-20</td>
<td>26417-26437</td>
<td>0.747166</td>
<td>0.639634</td>
<td>ACAAUUGGAUUGAACUUUG</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SARS-CoV-2-conserved-unstructured-21</td>
<td>12195-12210</td>
<td>0.745911</td>
<td>0.634404</td>
<td>UUAAPAAGUGAGAAGA</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SARS-CoV-2-conserved-unstructured-22</td>
<td>6757-6783</td>
<td>0.78979</td>
<td>0.633073</td>
<td>UUGUACUAUAUAUUGCUUACCUU</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SARS-CoV-2-conserved-unstructured-23</td>
<td>15236-15257</td>
<td>0.747024</td>
<td>0.630481</td>
<td>ACAACAGUUAUAAACUGUUAA</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SARS-CoV-2-conserved-unstructured-24</td>
<td>6225-6238</td>
<td>0.734349</td>
<td>0.628462</td>
<td>AUAACUCUAUUGU</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SARS-CoV-2-conserved-unstructured-25</td>
<td>9578-9598</td>
<td>0.825697</td>
<td>0.627812</td>
<td>UAUCUCGUAUAUACCUGUAC</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SARS-CoV-2-conserved-unstructured-26</td>
<td>21649-21662</td>
<td>0.820527</td>
<td>0.626762</td>
<td>UAUUCUUCACAC</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SARS-CoV-2-conserved-unstructured-27</td>
<td>23985-23998</td>
<td>0.799754</td>
<td>0.625346</td>
<td>AUCACUAAAAACCA</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SARS-CoV-2-conserved-unstructured-28</td>
<td>7161-7174</td>
<td>0.774316</td>
<td>0.62504</td>
<td>ACACCUCUCUCU</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SARS-CoV-2-conserved-unstructured-29</td>
<td>6010-6029</td>
<td>0.738649</td>
<td>0.624805</td>
<td>ACCCAACCAACCAUCCAA</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SARS-CoV-2-conserved-unstructured-30</td>
<td>6515-6529</td>
<td>0.81084</td>
<td>0.624018</td>
<td>UUAUAUAUACAGAA</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SARS-CoV-2-conserved-unstructured-31</td>
<td>18219-18232</td>
<td>0.749532</td>
<td>0.623586</td>
<td>UUAACUUAUUCUAC</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SARS-CoV-2-conserved-unstructured-32</td>
<td>11659-11681</td>
<td>0.819159</td>
<td>0.623162</td>
<td>UUUCUACACGGCUUACUAGAC</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SARS-CoV-2-conserved-unstructured-33</td>
<td>24778-24797</td>
<td>0.771131</td>
<td>0.622469</td>
<td>AAAGAUCUACACCCCUGCUC</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SARS-CoV-2-conserved-unstructured-34</td>
<td>21669-21683</td>
<td>0.788644</td>
<td>0.621176</td>
<td>UUUUAUACCCUGAC</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SARS-CoV-2-conserved-unstructured-35</td>
<td>6105-6122</td>
<td>0.777239</td>
<td>0.618997</td>
<td>AUAAAGAAAAACCUUAC</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SARS-CoV-2-conserved-unstructured-36</td>
<td>28436-28452</td>
<td>0.80827</td>
<td>0.618884</td>
<td>GCUUCACUCUACACUUG</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.27.012906. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder. It is made available under a CC-BY 4.0 International license.
Table 3: SARS-CoV-2-conserved-unstructured. Top unstructured regions (ranked by minimum unpaired probability over the interval, stretch of at least 15 nt) that overlap with conserved intervals from SARS-CoV-2 for at least 15 nt at a 97% sequence conservation cutoff. Sequence intervals are relative to the reference genome NC_045512.2.
Table 4: Sequences and secondary structure dot-brackets for key structured genome regions of SARS-CoV-2.
References


