RNA genome conservation and secondary structure in SARS-CoV-2 and SARS-related viruses: a first look

RAMYA RANGAN,[1] IVAN N. ZHELUDEV,[2] RACHEL J. HAGEY,[3] EDWARD A. PHAM,[3] HANNAH K. WAYMENT-STEELE,[4] JEFFREY S. GLENN,[3,5] and RHIJU DAS[1,2,6]

[1]Biophysics Program, Stanford University, Stanford, California 94305, USA

[2]Department of Biochemistry, Stanford University School of Medicine, Stanford, California 94305, USA

[3]Departments of Medicine (Division of Gastroenterology and Hepatology) and Microbiology & Immunology, Stanford School of Medicine, Stanford, California 94305, USA

[4]Department of Chemistry, Stanford University, Stanford, California 94305, USA

[5]Palo Alto Veterans Administration, Palo Alto, California 94304, USA

[6]Department of Physics, Stanford University, Stanford, California 94305, USA

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. Here we present a first look at RNA sequence conservation and structural propensities in the SARS-CoV-2 genome. Using sequence alignments spanning a range of betacoronaviruses, we rank genomic regions by RNA sequence conservation, identifying 79 regions of length at least 15 nt 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 subsequently reported from the 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 extended 5′ UTR, frameshifting stimulation element, and 3′ UTR. Lastly, we predict regions of the SARS-CoV-2 viral genome that 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 accessible by hybridization in primer-based diagnostic strategies.

Keywords: secondary structure; conservation; SARS-CoV-2; structurome; ncRNA

INTRODUCTION

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) has caused a rapidly expanding global pandemic, with the Coronavirus Disease 2019 (COVID-19) outbreak responsible at the time of writing for over 600,000 cases and 25,000 deaths (Rangan et al. 2020b). 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 (Fehr and Perlman 2015). 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 (Spurgers et al. 2008; Connelly et al. 2016; Bennett et al. 2019).

Conserved structured RNA regions have already been shown to play critical functional roles in the life cycles of

Corresponding authors: rhiju@stanford.edu, jsglenn@stanford.edu

Article is online at http://www.rnajournal.org/cgi/doi/10.1261/rna.076141.120. Freely available online through the *RNA* Open Access option.

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 (Yang and Leibowitz 2015). 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 (van den Born et al. 2005). Beyond the 5′ UTR, the frameshifting stimulation element (FSE) in the first protein-coding ORF (ORF1ab) includes a pseudoknot structure that is necessary for the ratiometric translation of ORF1a and ORF1ab from two overlapping reading frames via programmed ribosomal −1 frame-shifting (Plant and Dinman 2008). In the 3′ UTR, mutually exclusive RNA structures including the 3′ UTR pseudoknot control various stages of the RNA synthesis pathway (Stammler et al. 2011).

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 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. Many of our conclusions, first reported in the bioRxiv preprint server (Rangan et al. 2020b), have been confirmed in a later bioRxiv report (Andrews et al. 2020), which we discuss briefly.

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. To ensure reasonable numbers of sequences while still focusing on conservation and structure patterns most relevant to the current pandemic,

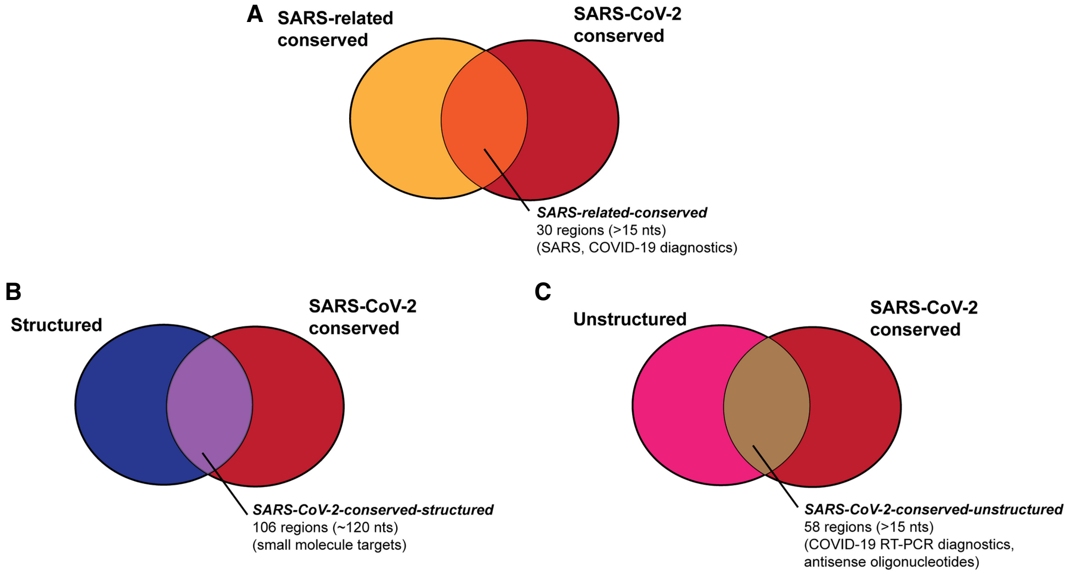

FIGURE 1. We aim to provide a series of genome regions in SARS-CoV-2 that are useful for a variety of diagnostic and therapeutic strategies, including regions that are (*A*) conserved in SARS-related betacoronaviruses and SARS-CoV-2 sequences (Table 1), (*B*) regions that are structured and conserved in SARS-CoV-2 sequences (Table 2), and (*C*) regions that are unstructured and conserved in SARS-CoV-2 sequences (Table 3).

we chose to analyze not all betacoronaviruses but a subgroup of SARS-related betacoronaviruses. These include SARS, SARS-CoV-2, and SARS-related bat coronaviruses, but not MERS, MHV, or other betacoronaviruses which have been classified into distinct subgroups based on different sequence and structure features in, for example, their 5′ UTR's (Chen and Olsthoorn 2010).

We carried out this analysis beginning with three different sequence alignments. Each captures a range of complete genome sequences across the SARS-related betacoronaviruses, but differ in the total number of sequences and in the redundancy of those sequences, as follows:

1. The first multiple sequence alignment (SARSr-MSA-1) was computed by aligning sequences curated by Ceraolo and Giorgi (2020), filtered by including only the reference genome sequence NC_0405512.2 (Wu et al. 2020) from the SARS-CoV-2 sequence set, removing the two MERS sequences, and leaving in all remaining betacoronavirus whole-genome sequences. This alignment captures a range of SARS-related bat coronavirus and SARS sequences but comprises only 11 sequences. These sequences correspond well to the SARS-related group defined in Gorbalenya et al. (Viruses and Coronaviridae Study Group of the International Committee on Taxonomy of 2020).
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, and removing mutually similar sequences with a 99% sequence conservation cutoff. With 180 sequences with at most 99% pairwise sequence 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 nt 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 nt or more enables the design

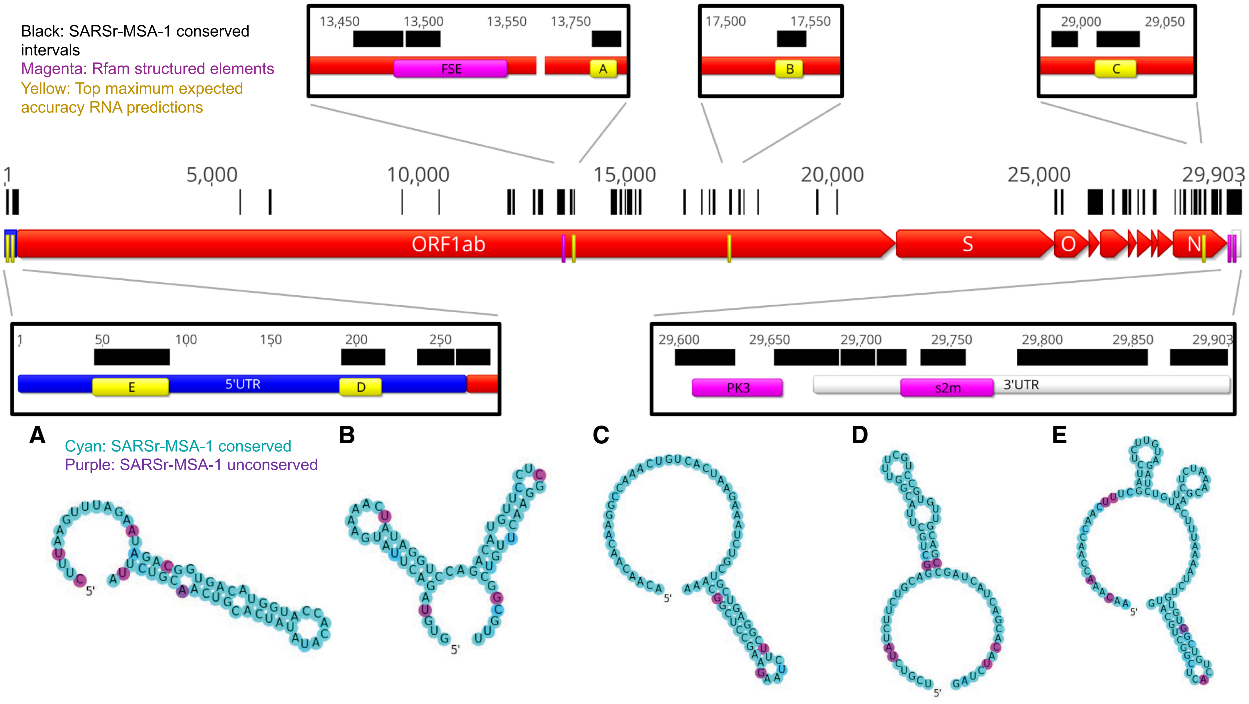

FIGURE 2. In black we annotate SARSr-MSA-1 conserved regions of the genome, superimposed on SARS-CoV-2 genome ORFs. We depict the top secondary structures as ranked by Matthews correlation coefficient that overlap with these conserved regions, ordered from *A* to *E*. Regions *A* to *E* are annotated 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 frameshifting stimulation element (FSE), the 3′ UTR pseudoknot (PK3), and the 3′ stem–loop II-like motif (s2m). Figures prepared in Geneious (Kearse et al. 2012) and draw_rna (https://github.com/DasLab/draw_rna).

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 artifacts.

In Figure 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, as expected based on prior work demonstrating sequence conservation surrounding structured RNA elements in these regions (Madhugiri et al. 2014), 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 (Kalvari et al. 2018) 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). Locations for the frameshifting stimulation element, 3′ UTR pseudoknot, and 3′ stem–loop II-like motif were confirmed using Infernal (Nawrocki and Eddy 2013), with all regions discovered at an $E < 10^{-4}$ threshold. We also found overlap between conserved stretches and additional 5′ UTR structures that have been established for previous coronaviruses, including the original SARS virus, including stem–loops 2–3 (SL2–3) and stem–loop 5 (SL5) (Yang et al. 2015). 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.

To further tighten this list of conserved sequences to ones most relevant to the current COVID-19 outbreak, we analyzed whether sequence regions conserved across SARS and bat coronaviruses remain conserved in the SARS-CoV-2 strains, most of which emerged after our analysis above (Fig. 1A). 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 nt in both cases: the first includes 103 NCBI sequences (SARS-CoV-2-MSA-1), and the second includes 739 sequences deposited to GISAID (Elbe and Buckland-Merrett 2017) (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 $<1 \times 10^{-5}$). 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 whole-genome sequences from the SARS-CoV-2 outbreak as of March 18, 2020 (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 Supplemental 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 isolates to date.

Predictions for structured regions in SARS-CoV-2

The intrinsic RNA structure of a conserved genome region is of interest in current medical research (Fig. 1B). On one hand, stable secondary structure domains are candidates for harboring stereotyped 3D RNA folds that present targets for small-molecule drug therapeutics. On the other hand, if an RNA region is sufficiently unstructured to allow binding by hybridization probes, antisense oligonucleotides may be used to disrupt these functional structures. Such unstructured stretches may also be more likely to be accessible to diagnostic strategies including standard RT-PCR assays (Bustin and Nolan 2004).

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 nt long, 100% sequence conservation) along with 20 nt flanking windows, we predicted maximum expected accuracy (MEA) secondary structures using CONTRAfold 2.0 (Do et al. 2006). 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 estimated Matthews correlation coefficient (MCC) from each construct's base-pairing probability matrix (Hamada et al. 2010). We note here that while MCC is often used in the RNA structure modeling literature to assess agreement of a prediction with a reference structure, we here use the metric to assess how tightly concentrated the ensemble of predicted secondary structures is to a single predicted secondary structure, the MEA structure. 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 Figure 2 regions A–E, we display the five conserved regions with the top maximum expected accuracy (MEA) secondary structures as ranked by the estimated MCC (all regions listed in Supplemental File 1). Regions D and E occurred within the 5′ UTR and correspond to known SARS-related virus stem–loops SL5a and SL2, respectively. Interestingly, region A is close to but does not overlap with the frameshifting stimulation element; it lies 200 nt downstream from the FSE and could perhaps be involved in a more elaborate structure, as has been described for human coronavirus 229E and other coronaviruses (Herold and Siddell 1993).

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 (Gruber et al. 2010), 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 ncRNAs 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 nt sliding by 40 nt, 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 117 windows with predicted structure ($P > 0.9$) (Fig. 3). These structured loci cover 46% of the SARS-CoV-2 genome, suggesting that the SARS-CoV-2 genome is highly structured, as has also been found in a recent in silico map of the SARS-CoV-2 RNA structurome (Andrews et al. 2020). We found that five canonical RNA structures (the frameshifting stimulation element, the 3′ UTR pseudoknot, the 3′ UTR hypervariable region, 5′ UTR SL2-3, and 5′ UTR SL5) were present in these 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 nt with RNAz loci. 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 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 nt 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 Supplemental File 1). Top-scoring structured windows from RNAz that overlap with conserved sequence regions in SARS-CoV-2-MSA-2 for at least 15 nt 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 Supplemental File 1.

We sought to further check structured windows reported by RNAz using orthogonal approaches. First, we explored using R-scape to make structure predictions with

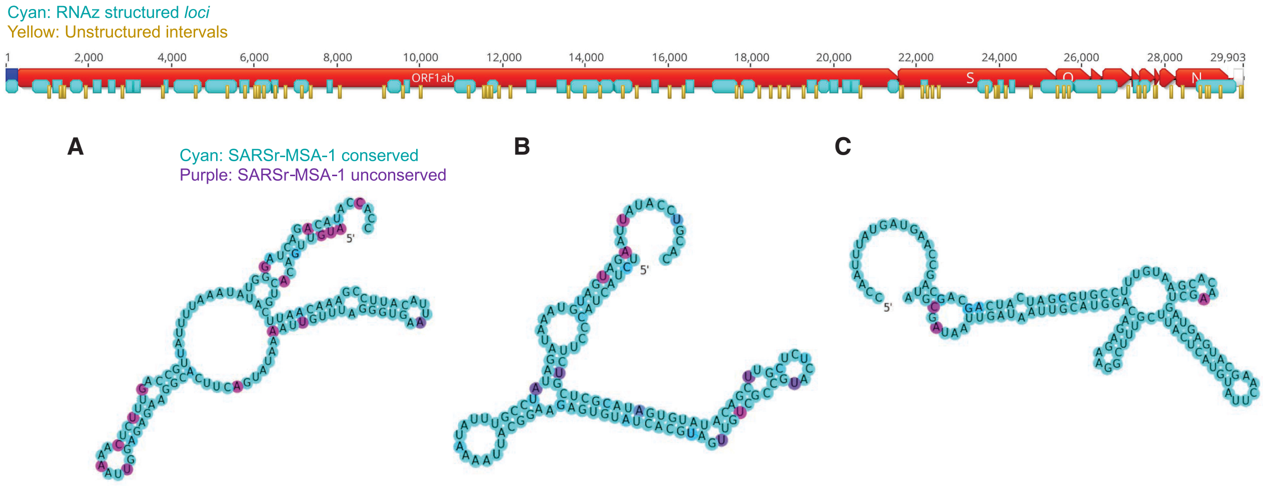
Cyan: RNAz structured *loci*
Yellow: Unstructured intervals
1
2,000
4,000
6,000
8,000
10,000
12,000
14,000
16,000
18,000
20,000
22,000
24,000
26,000
28,000
29,903
ORF1ab
S
O
N
A
Cyan: SARSr-MSA-1 conserved
Purple: SARSr-MSA-1 unconserved
B
C
5′
5′
5′

FIGURE 3. Structured (cyan) and unstructured (yellow) intervals on the genome ORFs for SARS-CoV-2, predicted from RNAz and CONTRAfold 2.0 analysis, respectively. *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). Figures prepared in Geneious (Kearse et al. 2012) and draw_rna (https://github.com/DasLab/draw_rna).

covariation signal in the sequence alignments (Rivas et al. 2020). However, we found that the SARSr-MSA-1 alignment had insufficient variation to detect conserved base pairs with covariation, lacking alignment power for all genomic windows (Rivas et al. 2020). Second, we compared windows predicted from RNAz above to those computed from ScanFold in a recent analysis (Andrews et al. 2020). We find that 90 of the 117 RNAz structured windows ($P > 0.9$ cutoff) are predicted ScanFold structured regions that meet a P-value threshold of 0.05. This overlap is statistically significant (P-value $<1 \times 10^{-5}$). Third, we validated structured window predictions with alifoldz, a program that calculates a z-score for an alignment window by comparing the window's consensus minimum free energy structure to that of random shuffled alignments. To mirror the RNAz analysis above, we scanned through windows of length 120 nt sliding by 40 nt. We chose a z-score cutoff of −2.69, which kept only 1% of windows when running alifoldz on all shuffled windows across the genome. This approach led to predicting 228 alifoldz structured windows, overlapping with 104 of the 117 RNAz structured windows ($P > 0.9$ cutoff). Again, this overlap is statistically significant (P-value $<1 \times 10^{-5}$). RNAz structured windows supported by alifoldz analysis are highlighted in Supplemental 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 (Fig. 1C). We scanned the SARS-CoV-2 reference genome in windows of length 120 nt sliding by 40 nt, 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 nt where every base has average base-pairing probability at most 0.4.

It is interesting to note that some structured 120 nt windows reported by RNAz include these unpaired stretches. A simple explanation for this observation is that such regions may encode for well-defined, conserved RNA structures that themselves harbor long unpaired loops to recruit proteins, distal RNA elements, or other molecular machinery.

Overall, we find that 58 of these unpaired stretches have at least 15 nt 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 Supplemental File 1).

As an orthogonal check for the unstructured intervals predicted using CONTRAfold 2.0 base-pairing probabilities, we used Vienna's RNAplfold to compute unpaired probabilities for each genome position. In general, we found that RNAplfold predicted lower unpaired probabilities than CONTRAfold 2.0, with only 10 intervals of length at least 15 nt having average base-pairing probability at most 0.4, in contrast with the 76 stretches predicted by CONTRAfold 2.0. Nevertheless, we found that 9 of the 10 intervals predicted by Vienna's RNAplfold overlap with unpaired intervals predicted from our CONTRAfold 2.0 analysis (regions listed in Supplemental 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 extended partially into the coding region of ORF1ab (Fig. 4A), frameshifting stimulation element (Fig. 4B), and 3′ UTR (Fig. 4C) for SARS-CoV-2. These models are built by analyzing homology with literature-annotated structures in related betacoronaviruses (Chen and Olsthoorn 2010; Wolfinger 2020) and by comparing to predicted secondary structure models from physics-based secondary structure prediction algorithms. The secondary structure for the 5′ UTR is the maximum expected accuracy (MEA) structure predicted by CONTRAfold 2.0. We additionally include computer-readable secondary structures in Table 4 and Supplemental 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 extended 5′ UTR includes five confident stem–loop structures (SL1–SL5), with structures verified by chemical mapping experiments in related coronaviruses (Chen and Olsthoorn 2010; Yang et al. 2015). SL1 and SL2 are conserved across betacoronaviruses, with SL2 having the highest sequence conservation across the 5′ UTR (Madhugiri et al. 2014). 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 (Li et al. 2008). SL2 has been shown to be critical for subgenomic RNA synthesis, with mutations in its conserved pentaloop retaining the production of genome-sized RNA, but not subgenomic RNA segments (Liu et al. 2007). 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" that gives rise to discontinuous transcription of subgenomic mRNAs (van den Born et al. 2005). SL4 contains a short upstream ORF, here labeled 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 (Madhugiri et al. 2014). SL5 has been implicated in a potential role in viral

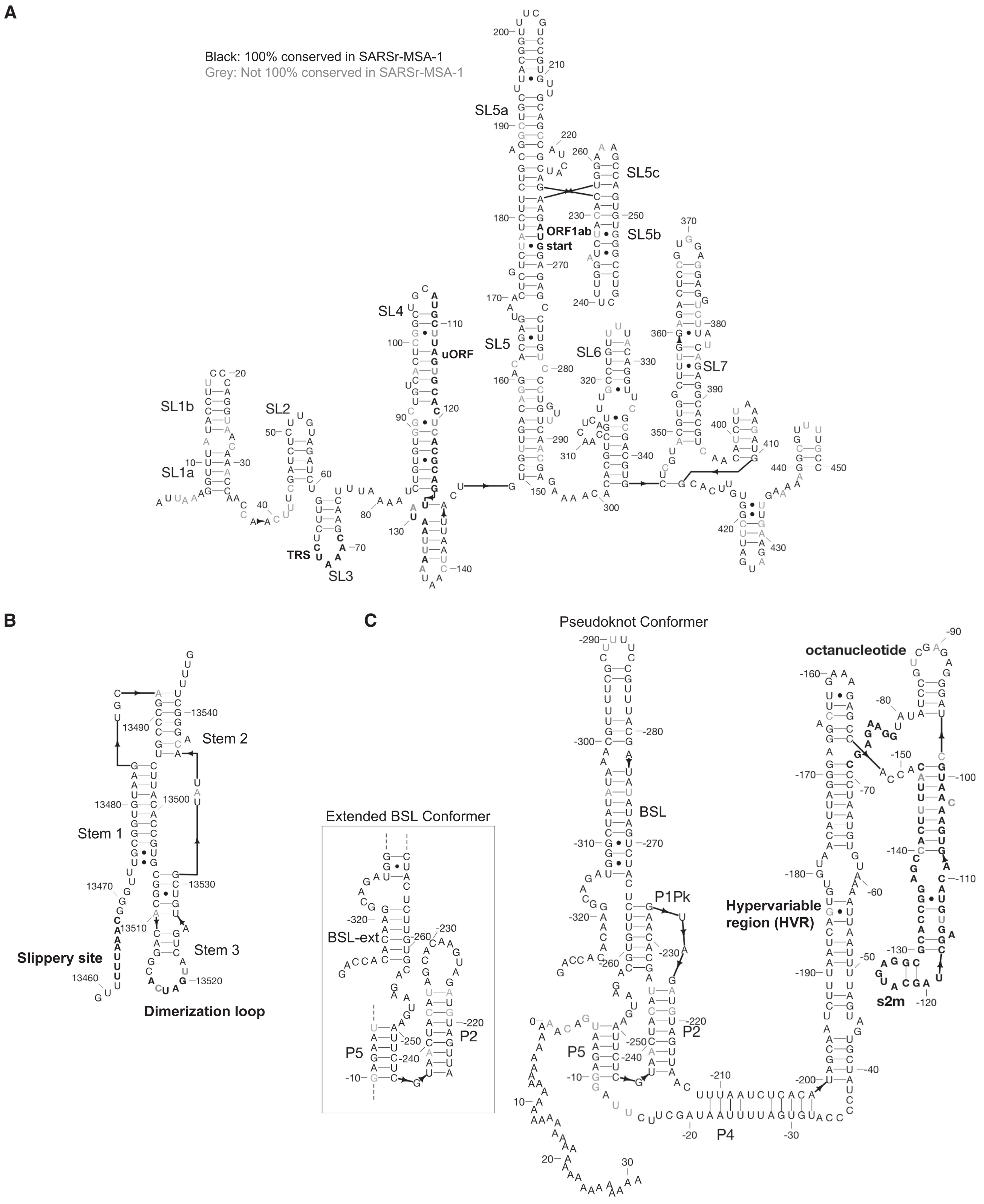

FIGURE 4. Secondary structure diagrams for (*A*) 5′ UTR, (*B*) frameshifting stimulation element, (*C*) 3′ UTR. Nucleotides are black if 100% conserved in the SARS, bat, and SARS-CoV-2 sequences in SARSr-MSA-1, and gray otherwise. Special labeled domains are in boldface. Structures are based primarily on manual identification of homology with literature coronavirus structure models. Note that numbering in *C* is relative to 3′ end of virus sequence. Figures prepared in RiboDraw (https://github.com/ribokit/RiboDraw).

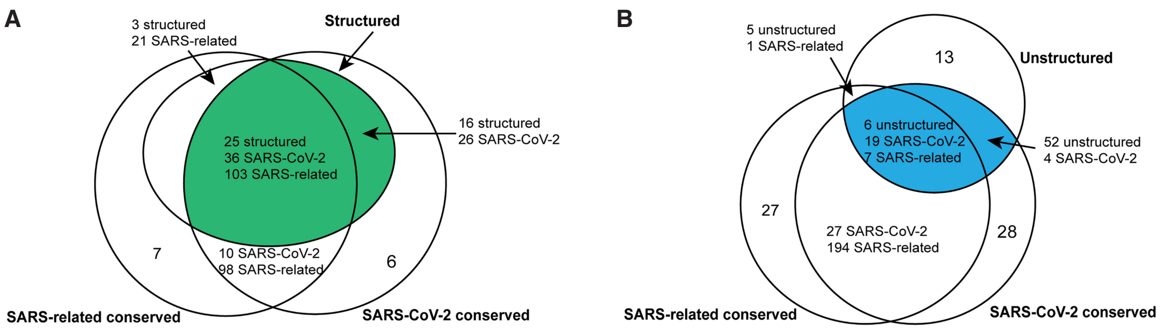

FIGURE 5. We depict the predicted number of structured, unstructured, and conserved intervals for a choice of sequence conservation cutoffs. The SARS-related conserved intervals are all regions of at least 15 nt with each position at least 90% conserved across an alignment of SARS, bat coronavirus, and SARS-CoV-2 sequences (SARSr-MSA-1). The SARS-CoV-2 intervals are regions of at least 15 nt 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 nt where all bases have base-pairing probability at most 0.4. All interval intersections are required to have at least 15 nt 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.

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 pentaloops on SL5a and SL5b, and a GNRA tetraloop on SL5c (Chen and Olsthoorn 2010). Prior DMS-probing data for Stem 5 in SARS-CoV aligned with the proposed SL5a,b,c structures (Chen and Olsthoorn 2010). Two additional stems (SL6 and SL7) are predicted in the MEA structure from CONTRAfold 2.0, but prior literature has not established whether such stems embedded in the coding region are functionally important across betacoronaviruses.

The frameshifting stimulation element (FSE) is located in ORF1ab and is involved in regulating a (−1) ribosomal frameshift event that is necessary for completing translation of ORF1ab. The FSE consists of a conserved pseudoknot structure that regulates the rate of ribosomal frameshifting at an upstream slippery site (Plant and Dinman 2008). 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 (Kelly and Dinman 2020).

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 (Goebel et al. 2004; van den Born et al. 2005). 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 (Goebel et al. 2007). 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 (Robertson et al. 2005). 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 (Jonassen et al. 1998).

DISCUSSION

Understanding the RNA structure of the SARS-CoV-2 genome can guide RNA-targeting interventions and diagnostic development. 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 previously established coronavirus structured elements that recur in the SARS-CoV-2 genome: the extended 5′ UTR, the frameshifting stimulation element, and the 3′ UTR. By filtering for sequences that have more than one of these properties, we have curated three sets of RNA genomic regions of potential interest for further structural analysis, which we have termed the SARS-related-conserved, SARS-CoV-2-conserved-structured, and SARS-CoV-2-conserved-unstructured sets. Figure 5 gives a more extensive presentation of how these sets overlap. Our hope is that these steps will provide useful starting points for efforts to develop antivirals and diagnostics that depend on targeting either structured or unstructured viral genomic regions.

The alignments used here act as an initial starting point for evaluating RNA sequence conservation in SARS-related viruses and in SARS-CoV-2 strains. Further conservation analysis may benefit from exploring the use of codon-correct viral genome alignments (Libin et al. 2019) or reference-guided alignments (Tzou et al. 2017). Additionally, alignments capturing a broader range of betacoronavirus sequences may enable the use of covariation signals for secondary structure prediction with tools like R-scape (Rivas et al. 2017).

The abundant RNA structures involved in the replication cycle of betacoronaviruses present ample opportunities for therapeutic development, but our analysis is not complete. First, while homology with prior structures annotated in betacoronaviruses lends some confidence to the 5′ UTR, frameshifting stimulation element, and 3′ UTR secondary structures presented here, there may still be some inaccuracies. For instance, SL6 and SL7 in the 5′ UTR are built based on computer modeling. As another example, the frameshifting stimulation element structure presented here differs in two base pairs compared to that presented by Kelly and Dinman (2020). While a recent analysis of the SARS-CoV-2 structurome conducted with ScanFold (Andrews et al. 2018) supports various base pairs in the homology models for the 5′ UTR, FSE, and 3′ UTR presented here, base pairs in the SL5 basal stem of the 5′ UTR and in the hyper-variable region of the 3′ UTR are not predicted by ScanFold (Andrews et al. 2020). Some of these base pairs would not be found by ScanFold because they involve pairing of nucleotides that extend beyond the ScanFold window size used (120 nt), as is the case for the SL5 basal stem in the extended 5′ UTR and the basal portion of the 3′ UTR hyper-variable region stem. Some other base pairs involved in pseudoknots would not be found by ScanFold. However, numerous base pairs in the hyper-variable region model presented here (29695–29809 in numbering based on reference genome NC_0405512.2 [Wu et al. 2020], or −117 to −63 in numbering corresponding to Fig. 4C) that could be identified by ScanFold windowed analysis were not found by ScanFold; it is possible that RNA in these regions adopts a heterogeneous set of secondary structures or a tertiary structure stabilized by noncanonical pairs that are not modeled in ScanFold. Additional biochemical and genetic verification, particularly through compensatory mutagenesis, will be needed to further assess these structures.

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 coevolving with specific mutations in the 3′ UTR (Guan et al. 2012; Madhugiri et al. 2014). Such long-range RNA–RNA interactions would be missed in our analyses above which have focused on shorter windows. In related coronaviruses, RNA structures that act as packaging signals have been identified in genomic ORFs (Madhugiri et al. 2014). Such packaging signals may reside in regions identified here from our RNAz computational analysis or may have been missed. Last, it seems likely that the reverse complement of the RNA genome harbors important functional elements which we have not analyzed in detail in this study.

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 (Kappel et al. 2019; Zhang et al. 2019). 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 secondary structures predicted here present starting points for 3D modeling of RNA-only structures in various regions, including the 5′ UTR SL5, the frameshifting stimulation element, the 3′ UTR pseudoknot, and the 3′ stem–loop II-like motif, and this modeling has begun (Rangan et al. 2020a). Furthermore, these regions and novel predicted structures 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.

MATERIALS AND METHODS

Open reading frame (ORF) annotation

The ORFs included in Geneious (Kearse et al. 2012) genome maps (Figs. 2, 3) were obtained from protein annotations deposited in NCBI for the reference genome NC_0405512.2 (Wu et al. 2020). We exclude ORF10, as recent reports have found no evidence for ORF10 expression in SARS-CoV-2 (Kim et al. 2020).

Conservation analysis

Three alignments for SARS-related viruses were prepared:

1. SARSr-MSA-1 was generated by realigning the sequences curated by Ceraolo and Giorgi (Ceraolo and Giorgi 2020) with MUSCLE (Edgar 2004) using default alignment settings, excluding all nonreference genome copies of the SARS-CoV-2 sequence and excluding MERS sequences JX869059.2 and KT368829.1.
2. SARSr-MSA-2 was generated by downloading the MSA provided by BLAST for the top 100 complete genome sequences closest to the SARS-CoV-2 reference genome.
3. SARSr-MSA-3 was generated by obtaining all complete betacoronavirus genome sequences available from the NCBI database, removing mutually similar sequences using a 99% cutoff

TABLE 1. SARS-related-conserved

Name	Interval	Sequence	Conservation in SARS-CoV-2
SARS-related-conserved-1	14060–14075	UAGAUAAUCAAGAUCU	0.995896
SARS-related-conserved-2	15838–15857	UGGACUGAGACUGACCUUAC	0.995896
SARS-related-conserved-3	28554–28569	UUCGUGGUGGUGACGG	0.995868
SARS-related-conserved-4	28513–28546	AGAUGACCAAAUUGGCUACUACCGAAGAGCUACC	0.995868
SARS-related-conserved-5	16153–16169	GUUAUGCUUACUAAUGA	0.994528
SARS-related-conserved-6	27183–27212	GUACAGUAAGUGACAACAGAUGUUUCAUCU	0.99449
SARS-related-conserved-7	27165–27181	GACAAUAUUGCUUUGCU	0.99449
SARS-related-conserved-8	25511–25530	CACUCCCUUUCGGAUGGCUU	0.99449
SARS-related-conserved-9	25393–25409	AUGGAUUUGUUUAUGAG	0.99449
SARS-related-conserved-10	12905–12924	AGGUUUGUUACAGACACACC	0.99316
SARS-related-conserved-11	13346–13361	GUGGGUUUUACACUUA	0.99316
SARS-related-conserved-12	15496–15518	ACAACUGCUUAUGCUAAUAGUGU	0.99316
SARS-related-conserved-13	28799–28818	AGCAGAGGCGGCAGUCAAGC	0.993113
SARS-related-conserved-14	27457–27473	GAGUGUGUUAGAGGUAC	0.993113
SARS-related-conserved-15	25547–25562	UUCUUGCUGUUUUUCA	0.993113
SARS-related-conserved-16	17089–17105	GGUACUGGUAAGAGUCA	0.991792
SARS-related-conserved-17	17956–17975	UGCAUAAUGUCUGAUAGAGA	0.991792
SARS-related-conserved-18	18034–18050	UUACAAGCUGAAAAUGU	0.991792
SARS-related-conserved-19	704–723	GACGAGCUUGGCACUGAUCC	0.991792
SARS-related-conserved-20	25376–25392	UACACAUAAACGAACUU	0.991736
SARS-related-conserved-21	10406–10422	UACAAUGGUUCACCAUC	0.990437
SARS-related-conserved-22	16364–16388	UGUCUGUUAAUCCGUAUGUUUGCAA	0.990424
SARS-related-conserved-23	15622–15644	UAUGAGUGUCUCUAUAGAAAUAG	0.990424
SARS-related-conserved-24	15349–15367	GUUCUUGCUCGCAAACAUA	0.990424
SARS-related-conserved-25	15301–15323	AAAUGUGAUAGAGCCAUGCCUAA	0.990424
SARS-related-conserved-26	14077–14099	AAUGGUAACUGGUAUGAUUUCGG	0.990424
SARS-related-conserved-27	741–756	AAAACUGGAACACUAA	0.990424
SARS-related-conserved-28	25106–25128	GAAAUUGACCGCCUCAAUGAGGU	0.990358
SARS-related-conserved-29	26232–26267	UGAGUACGAACUUAUGUACUCAUUCGUUUCGGAAGA	0.990358
SARS-related-conserved-30	28270–28293	UAAAAUGUCUGAUAAUGGACCCCA	0.990358

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 nt, 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.

with CD-HIT-EST (Li et al. 2001), and computing an MSA with Clustal Omega (Sievers et al. 2011) using default settings.

Two alignments for SARS-CoV-2 sequences were prepared:

1. SARS-CoV-2-MSA-1 was generated by downloading the MSA provided by NCBI for the 103 whole-genome SARS-CoV-2 sequences deposited as of 03-18-20.
2. SARS-CoV-2-MSA-2 was generated from 739 GISAID (Elbe and Buckland-Merrett 2017) sequences, including all sequences described in the Nextstrain project metadata file as of 03-18-20 (https://github.com/nextstrain/ncov). Sequences were aligned using MAFFT (Katoh and Frith 2012) with the –add flag to add GISAID sequences to seed alignment SARS-CoV-2-MSA-1, and the sequences in SARS-CoV-2-MSA-1 were subsequently removed to avoid duplicates.

All alignments are included in the associated GitHub repository: https://github.com/DasLab/SARSCoV2_Secstruct_Cons.

Analysis of structured elements

To identify regions in the SARS-CoV-2 reference genome NC_0405512.2 (Wu et al. 2020) that were matches to Rfam (Kalvari et al. 2018) families, we used Infernal (Nawrocki and Eddy 2013) to build covariance models from Rfam families RR0164, RR0165, and RR0507 with cmbuild, and we ran cmscan to find hits with an $E < 1 \times 10^{-4}$ threshold.

MEA structures were computed using CONTRAfold 2.0 (Do et al. 2006) for 20 nt flanking windows around conserved intervals in SARS-CoV-2. To estimate the Matthews correlation coefficient of these single-structure predictions, we computed the pseudo-expected Matthews correlation coefficient as previously described (Hamada et al. 2010). Base-pairing probability matrices

TABLE 2. SARS-CoV-2-conserved-structured

Name	Sequence interval	Sequence	Secondary structure	z-score	P value
SARS-CoV-2-conserved-structured-1	918–1037	ACUGGACUUUAUUGACACUAAGAGGGGUGUAUAC UGCUGCCGUGAACAUGAGCAUGAAAUUGCUUGGU ACACGGAACGUUCUGAAAAGAGCUAUGAAUUGCA GACACCUUUUGAAAUUAA	....(((((((((... ((((..(((((.((. ((((((......)))))))).))))))...(((((....))))) ....)))))))))))))........	−4.38	0.999
SARS-CoV-2-conserved-structured-2	14206–14325	AUGUUGACACUGACUUAACAAAGCCUUACAUUAA GUGGGAUUUGUUAAAAUAUGACUUCACGGAAGAG AGGUUAAAACUCUUUGACCGUUAUUUUAAAUAUU GGGAUCAGACAUACCACC	(((((((..(..(.((((((((..(((((.....))))).)))))))) ...........((((.((((((........))))))..)))) ....)..)..))).))))......	−1.99	0.999
SARS-CoV-2-conserved-structured-3	17125–17244	UCUACUACCCUUCUGCUCGCAUAGUGUAUACAGC UUGCUCUCAUGCCGCUGUUGAUGCACUAUGUGAG AAGGCAUUAAAAUAUUUGCCUAUAGAUAAAUGUA GUAGAAUUAUACCUGCAC	((((((((...((((((((((((((((((((..((......)).)))))..))))))))))))).(((((.......... .))))).)))) .....)))))))).............	−6.85	0.999
SARS-CoV-2-conserved-structured-4	7011–7130	UUUAGGUGUUUUAAUGUCUAAUUUAGGCAUGCCU UCUUACUGUACUGGUUACAGAGAAGGCUAUUUGA ACUCUACUAAUGUCACUAUUGCAACCUACUGUAC UGGUUCUAUACCUUGUAG	..((((((((..((((((((...)))))))(((((((.. (((((.....))))))))))) ....))..)).)))))...(((. (((.....)))..))).	−1.36	0.999
SARS-CoV-2-conserved-structured-5	26135–26254	AAUCGACGGUUCAUCCGGAGUUGUUAAUCCAGUA AUGGAACCAAUUUAUGAUGAACCGACGACGACUA CUAGCGUGCCUUUGUAAGCACAAGCUGAUGAGUA CGAACUUAUGUACUCAUU	..(((.((((((((..((((((....((((....))))..))))))..))))))))).))).......((((((((.........))))..)))) (((((((((......))))))))).	−5.44	0.999
SARS-CoV-2-conserved-structured-6	26175–26294	CCAAUUUAUGAUGAACCGACGACGACUACUAGCG UGCCUUUGUAAGCACAAGCUGAUGAGUACGAACU UAUGUACUCAUUCGUUUCGGAAGAGACAGGUACG UUAAUAGUUAAUAGCGUA	...................(((..(((((.((((((((((..... (((....)))((((((((((......)))))))))). (((((....))))))))))))))).))))).....))).	−4.44	0.999
SARS-CoV-2-conserved-structured-7	6251–6370	AUGCAACUAAUAAAGCCACGUAUAAACCAAAUAC CUGGUGUAUACGUUGUCUUUGGAGCACAAAACCA GUUGAAACAUCAAAUUCGUUUGAUGUACUGAAGU CAGAGGACGCGCAGGGAA	.(((.......((((((.(((((((.((((......)))).))))))).).)))))..((........((((....(((((((((....)))))))))))).. (((....)))))))).....	−2.52	0.999
SARS-CoV-2-conserved-structured-8	40–157	UUUCGAUCUCUUGUAGAUCUGUUCUCUAAACGAA CUUUAAAAUCUGUGUGGCUGUCACUCGGCUGCAU GCUUAGUGCACUCACGCAGUAUAAUUAAUAACUA AUUACUGUCGUUGACA	....(((((.....)))))..........(((((.......... ((((((((..((.((((.(((.....))).))))))..)))))))).. ((((((.....))))))...))))))....	−2.91	0.998
SARS-CoV-2-conserved-structured-9	26491–26607	AGUUUUUCUGUUUGGAACUUUAAUUUUAGCCAUG GCAGAUUCCAACGGUACUAUUACCGUUGAAGAGC UUAAAAAGCUCCUUGAACAAUGGAACCUAGUAAU AGGUUUCCUAUUCCU	.((((.(((((((((..((.........))))).)))))))... ((((((((....))))))))..(((((.....)))))...))))... ((((((((....)))))))).........	−3.52	0.998
SARS-CoV-2-conserved-structured-10	25895–26014	CAUUACUUCAGGUGAUGGCACAACAAGUCCUAUU UCUGAACAUGACUACCAGAUUGGUGGUUAUACUG AAAAAUGGGAAUCUGGAGUAAAAGACUGUGUUGU AUUACACAGUUACUUCAC	..(((((((((((..((.......))..((((((((.(.(..((((((((((......)))))))))..).).))))))))))))))))))..(((((((((.......))))))))........	−4.4	0.998
SARS-CoV-2-conserved-structured-11	5211–5330	AUUAAAUCACACUAAAAAGUGGAAAUACCCACAA GUUAAUGGUUUAACUUCUAUUAAAUGGGCAGAUA ACAACUGUUAUCUUGCCACUGCAUUGUUAACACU CCAACAAAUAGAGUUGAA	..........((.(((...((((......))))...))).)).(((((((. ((((((.....(((((((((((....)))))).))))........ ((((.........))))))))))))))))	−2.76	0.998
SARS-CoV-2-conserved-structured-12	26215–26334	UGUAAGCACAAGCUGAUGAGUACGAACUUAUGUA CUCAUUCGUUUCGGAAGAGACAGGUACGUUAAUA GUUAAUAGCGUACUUCUUUUUCUUGCUUUCGUGG UAUUCUUGCUAGUUACAC	((((((((.((((.((((((((((......))))))))))).)))).((((((((.. ((((((((((.........))))))))))))))))).)))..... ((((((....))))).))))).	−3.49	0.998
SARS-CoV-2-conserved-structured-13	6211–6330	GCUAAAUUGUUACAUAAACCUAUUGUUUGGCAUG UUAACAAUGCAACUAAUAAAGCCACGUAUAAACC AAAUACCUGGUGUAUACGUUGUCUUUGGAGCACA AAACCAGUUGAAACAUCA	(((..((((((((((...((.........)).))).)))))))........((((((.(((((((.((((......)))).))))))).).))))).)))	−1.03	0.998

(*Continued*)

TABLE 2. *(Continued)*

Name	Sequence interval	Sequence	Secondary structure	z-score	P value
SARS-CoV-2-conserved-structured-14	16445–16564	UUAUUGUAAAUCACAUAAACCACCCAUUAGUUUUCCAUUGUGUGCUAAUGGACAAGUUUUUGGUUUAUAUAAAAAUACAUGUGUUGGUAGCGAUAAUGUUAC UGACUUUAAUGCAAUUGC	....((((((((...((((...((((((((............))))))).....)))).))))))))........(((..((..((((((......))))..))....))).......	−1.17	0.997
SARS-CoV-2-conserved-structured-15	7051–7170	UGUACUGGUUACAGAGAAGGCUAUUUGAACUCUACUAAUGUCACUAUUGCAACCUACUGUACUGGUUCUAUACCUUGUAGUGUUUGUCUUAGUGGUUUAGAU UCUUUAGACACCUAUCCU	(((.(((....))).(((((..((((((((...((....))(((((..(((((((((......(((.....)))..)))).).))))..))))))))))))))))))..)))........	0.99	0.996
SARS-CoV-2-conserved-structured-16	158–277	GGACACGAGUAACUCGUCUAUCUUCUGCAGGCUGCUUACGGUUUCGUCCGUGUUGCAGCCGAUCAUCAGCACAUCUAGGUUUCGUCCGGGUGUGACCGAAA GGUAAGAUGGAGAGCCUUG	.....((((...(((.((((((((.....((((((.((((((......)))))..))))))..........(((((((.((......))))))))(((....))))))))))))).))))	−2.87	0.996
SARS-CoV-2-conserved-structured-17	23977–24096	UUACCAGAUCCAUCAAAACCAAGCAAGAGGUCAUUUAUUGAAGAUCUACUUUUCAACAAAGUGACACUUGCAGAUGCUGGCUUCAUCAAACAAUAUGGUGAU UGCCUUGGUGAUAUUGCU	(((((((...............(((((..(((((((.((((((((.....)))))))..))))))).))))).......(((.((((((........))))))..))))))))))........	−2.82	0.994
SARS-CoV-2-conserved-structured-18	25855–25974	ACUAUUGUAUACCUUACAAUAGUGUAACUUCUUCAAUUGUCAUUACUUCAGGUGAUGGCACAACAAGUCCUAUUUCUGAACAUGACUACCAGAUUGGUGGUU AUACUGAAAAAUGGGAAU	((((((((((.....))))))))))..............((((((((((.....))))))))........((((((((.(.(..((((((((((.....))))))))))..).).))))))))..	−4.25	0.994
SARS-CoV-2-conserved-structured-19	2943–3062	GGGCAUUGAUUUAGAUGAGUGGAGUAUGGCUACAUACUACUUAUUUGAUGAGUCUGGUGAGUUUAAAUUGGCUUCACAUAUGUAUUGUUCUUUCUACCCUC CAGAUGAGGAUGAAGAAGA	..((((((..((((((((((..((((((.....))))))))))))))))((((.(..((.........))..).))))))).))))....((((((.....((((.....))))..)))))..	−0.32	0.994
SARS-CoV-2-conserved-structured-20	14726–14845	UAAGGAAGGAAGUUCUGUUGAAUUAAAACACUUCUUCUUUGCUCAGGAUGGUAAUGCUGCUAUCAGCGAUUAUGACUACUAUCGUUAUAAUCUACCAACAAU GUGUGAUAUCAGACAACU	..((.((((((((...((((........))))...))))))).)) .....(((((..((((.....))))((((((((((........))))))))))))).....(((.((.....)).)))....	−3.02	0.994
SARS-CoV-2-conserved-structured-21	26375–26490	CAAUAUUGUUAACGUGAGUCUUGUAAAACCUUCUUUUUACGUUUACUCUCGUGUUAAAAAUCUGAAUUCUUCUAGAGUUCCUGAUCUUCUGGUCUAAACGAA CUAAAUAUUAUAUU	.((((((.....(((.((((..((((((((.......))))))))...)))).)))(((.........((((((.....))))))).. ((((.....))))..))).......))))))......	−2.73	0.994
SARS-CoV-2-conserved-structured-22	21365–21484	AAUUCAGUUGUCUUCCUAUUCUUUAUUUGACAUGAGUAAAUUUCCCCUUAAAUUAAGGGGUACUGCUGUUAUGUCUUUAAAAGAAGGUCAAAUCAAUGAUAU GAUUUUAUCUCUUCUUAG	...................((((((.... ((((((((((......((((((((...))))))))....))))..)))))) ....))))))(((.((((((.......))))))).)))...........	−2.27	0.994
SARS-CoV-2-conserved-structured-23	2158–2277	GAAGAGAAGUUUAAGGAAGGUGUAGAGUUUCUUAGAGACGGUUGGGAAAUUGUUAAAUUUAUCUCAACCUGUGCUUGUGAAAUUGUCGGUGGACAAAUUGU CACCUGUGCAAAGGAAAUU	.........((((.((((((.(.....).))))))..))))(((((((((((((.....))))..))))))))((..(..((((..(((((.....))))) .....))))..)..))..........	−1.62	0.993
SARS-CoV-2-conserved-structured-24	4174–4293	AAAAAGGCUGGUGGCACUACUGAAAUGCUAGCGAAAGCUUUGAGAAAAGUGCCAACAGACAAUUAUAUAACCACUUACCCGGGUCAGGGUUUAAAUGGUUAC ACUGUAGAGGAGGCAAAG	......(.(((.(((((((.((.((......(((.....))))).))...))))))).))).)..(((((((((((...((((.......))))......))))))...)))))..........	−1.33	0.993
SARS-CoV-2-conserved-structured-25	26528–26647	GAUUCCAACGGUACUAUUACCGUUGAAGAGCUUAAAAAGCUCCUUGAACAAUGGAACCUAGUAAUAGGUUUCCUAUUCCUUACAUGGAUUUGUCUUCUACAA UUUGCCUAUGCCAACAGG	......((((((((.....)))))))..(((((......)))))((((.......((((((((.....))))))))...........(((((..((((......)))).....)).)))......))))	−3.29	0.993
SARS-CoV-2-conserved-structured-26	26255–26374	CGUUUCGGAAGAGACAGGUACGUUAAUAGUUAAUAGCGUACUUCUUUUUCUUGCUUUCGUGGUAUUCUUGCUAGUUACACUAGCCAUCCUUACUGCGCUUCG AUUGUGUGCGUACUGCUG	.(((((.....))))).((((((..((((((...((((((..........((((......))))......((((((......)))))..........))))))..))))))..))))))).....	−2.31	0.992
SARS-CoV-2-conserved-structured-27	5651–5770	AUCUAGUACAACAGGAGUCACCUUUUGUUAUGAUGUCAGCACCACCUGCUCAGUAUGAACUUAAGCAUGGUACAUUUACUUGUGCUAGUGAGUACACUGGUA AUUACCAGUGUGGUCACU	...(((((((((((((.....))))))))..(((((.....((((..((((.(((.....)))..))))))))))))......))))))((((.((((((((((.....)))))))))).)))).	−2.77	0.992

SARS-CoV-2-conserved-structured-28	7131–7250	UGUUUGUCUUAGUGGUUUAGAUUCUUUAGACACC UAUCCUUCUUUAGAAACUAUACAAAUUACCAUUU CAUCUUUUAAAUGGGAUUUAACUGCUUUUGGCUU AGUUGCAGAGUGGUUUUU	.((((((..((((.(((((((...))))))).. (((.........)))..)))).))))))..((((((.........)))))) .....(((..(((...((......)).)))..)))...	−0.12	0.992
SARS-CoV-2-conserved-structured-29	10970–11089	AAAGUGCAGUGAAAAGAACAAUCAAGGGUACACA CCACUGGUUGUUACUCACAAUUUUGACUUCACUU UUAGUUUUAGUCCAGAGUACUCAAUGGUCUUUGU UCUUUUUUUUGUAUGAAA	...((((((.((((((((((.((..(((....)))..))((((.(((((... (((..((((.........))))..)))...)))))..)))) .....))))))))))).))))))....	−1.44	0.992
SARS-CoV-2-conserved-structured-30	28838–28957	GUAGUCGCAACAGUUCAAGAAAUUCAACUCCAGG CAGCAGUAGGGGAACUUCUCCUGCUAGAAUGGCU GGCAAUGGCGGUGAUGCUGCUCUUGCUUUGCUG CUGCUUGACAGAUUGAACC	.(((((.....((((.((....)).)))).(((((((((((((((.....))))))))......(((.(((((. (((((.....))))).)))))..))))))))))...)))))....	−2.64	0.992
SARS-CoV-2-conserved-structured-31	26095–26214	AAAUUGUUGAUGAGCCUGAAGAACAUGUCCAAAU UCACACAAUCGACGGUUCAUCCGGAGUUGUUAAU CCAGUAAUGGAACCAAUUUAUGAUGAACCGACGA CGACUACUAGCGUGCCUU	.....((((.((..(.((.....)).)..))...........(((. ((((((((..((((((....((((....))))..))))))..)))))))).))))))))..............	−1.91	0.992
SARS-CoV-2-conserved-structured-32	2903–3022	UAAAAACUUUGCAACCAGUAUCUGAAUUACUUAC ACCACUGGGCAUUGAUUUAGAUGAGUGGAGUAU GGCUACAUACUACUUAUUUGAUGAGUCUGGUGAG UUUAAAUUGGCUUCACAUA	.............((((((.......)))(((.(((((..(((.((... ((((((((((..((((((....)))))))))))))))).)))))..))).)).)))..)))	−0.97	0.991
SARS-CoV-2-conserved-structured-33	29550–29669	AAGGCAGAUGGGCUAUAUAAACGUUUUCGCUUUU CCGUUUACGAUAUAUAGUCUACUCUUGUGCAGAA UGAAUUCUCGUAACUACAUAGCACAAGUAGAUGU AGUUAACUUUAAUCUCAC	.(((.((.((((((((((..(((...((.......))...))).))))))))))))))))(((.(((.((((......((((((((. ((....))..)))))))))..)))).))))))	−1.83	0.990
SARS-CoV-2-conserved-structured-34	80–197	AAUCUGUGUGGCUGUCACUCGGCUGCAUGCUUAG UGCACUCACGCAGUAUAAUUAAUAACUAAUUACU GUCGUUGACAGGACACGAGUAACUCGUCUAUCUU CUGCAGGCUGCUUACG	...(((((((..((.((((.(((.....))).))))))..)))))))..((((((.....))))))..((.(((..(((((.(((((...))))) .....)))))..))).)).....	−0.8	0.990
SARS-CoV-2-conserved-structured-35	13886–14005	CAAUUGUUGUGAUGAUGAUUAUUUCAAUAAAAAG GACUGGUAUGAUUUUGUAGAAAACCCAGAUAUAU UACGCGUAUACGCCAACUUAGGUGAACGUGUACG CCAAGCUUUGUUAAAAAC	...((((((.(((((...))))).))))))((((.. ((((...................))))..........(((((((((... ((....))...)))))))))....))))..........	−0.37	0.989
SARS-CoV-2-conserved-structured-36	120–237	CACGCAGUAUAAUUAAUAACUAAUUACUGUCGUU GACAGGACACGAGUAACUCGUCUAUCUUCUGCAG GCUGCUUACGGUUUCGUCCGUGUUGCAGCCGAUC AUCAGCACAUCUAGGU	...((((((..((((.....)))))))))).(((((((((.(((((...))))) .....)))))..((((((.(((((.......)))))..)))))) .....))))))...........	−2.33	0.989
SARS-CoV-2-conserved-structured-37	23697–23816	UGCCAUACCCACAAAUUUUACUAUUAGUGUUACC ACAGAAAUUCUACCAGUGUCUAUGACCAAGACAU CAGUAGAUUGUACAAUGUACAUUUGUGGUGAUUC AACUGAAUGCAGCAAUCU	.........................(((((((((((((((...(((((.. (((((((........))))))..))))). ((((((...))))))))))))))) ...)))))............	−2.64	0.989
SARS-CoV-2-conserved-structured-38	25655–25774	AACAGUUUACUCACACCUUUUGCUCGUUGCUGCU GGCCUUGAAGCCCCUUUUCUCUAUCUUUAUGCUU UAGUCUACUUCUUGCAGAGUAUAAACUUUGUAAG AAUAAUAAUGAGGCUUUG	..((((.................((.....)).)))) ((((((((((.....................)))))........(((((((((((((....)))))))))))).........)))))....	−1.15	0.989
SARS-CoV-2-conserved-structured-39	15606-1–5724	UUACAACACAGACUUUAUGAGUGUCUCUAUAGAA AUAGAGAUGUUGACACAGACUUUGUGAAUGAGUU UUACGCAUAUUUGCGUAAACAUUUCUCAAUGAUG AUACUCUCUGACGAUGC	.........((((.......(((((((..(((.((....((((((((.. (((((...)))))..........((((((....))))))))))))))))).)))..))))))))))........	−1.81	0.988
SARS-CoV-2-conserved-structured-40	14526–14645	AGCUCUAGACUUAGUUUUAAGGAAUUACUUGUGU AUGCUGCUGACCCUGCUAUGCACGCUGCUUCUGG UAAUCUAUUACUAGAUAAACGCACUACGUGCUUU UCAGUAGCUGCACUUACU	...........((((((....))))))...(((((.(((((((((..........(((((.(((.((((((((....)))))))) .....)))...)))))...)))))))))))).....	−1.64	0.986
SARS-CoV-2-conserved-structured-41	3808–3897	CUCUAUGACAAACUUGUUUCAAGCUUUUUGGAAA UGAAGAGUGAAAAGCAAGUUGAACAAAAGAUCGC UGAGAUUCCUAAAGAGGAAGUU	(((..(((..(((((((((...(((((((......)))))))...))))))))) ....)))..))).(((((....)))))...	−1.32	0.983
SARS-CoV-2-conserved-structured-42	25935–26054	CAUGACUACCAGAUUGGUGGUUAUACUGAAAAAU GGGAAUCUGGAGUAAAAGACUGUGUUGUAUUACA CAGUUACUUCACUUCAGACUAUUACCAGCUGUAC UCAACUCAAUUGAGUACA	.((((((((((.....))))))))).(((...(((((...((((((((((.. (((((((((.......))))))))..)).)))))))))))))..))).. ((((((((.....)))))))).	−4.97	0.983

(Continued)

TABLE 2. *(Continued)*

Name	Sequence interval	Sequence	Secondary structure	z-score	P value
SARS-CoV-2-conserved-structured-43	15645–15764	GAUGUUGACACAGACUUUGUGAAUGAGUUUUACGCAUAUUUGCGUAAACAUUUCUCAAUGAUGAUACUCUCUGACGAUGCUGUUGUGUGUUUCAAUAGCACU UAUGCAUCUCAAGGUCUA	(((.((((..((((..(..(.(.((((.((((((((....)))))))).....)))).).)..) .....))))..(((((...((.(((((.....))))).)).)))))))))).)))..	−2.07	0.983
SARS-CoV-2-conserved-structured-44	2518–2625	UUAACAGAGGAAGUUGUCUUGAAAACUGGUGAUUUACAACCAUUAGAACAACCUACUAGUGAAGCUGUUGAAGCUCCAUUGGUUGGUACACCAGUUUGUAUU AACGGG	............((((......((((((((.((((........))))...(((.(((((((.((((.....)))).)))))).))).)))))))))....))))...	−1.93	0.982
SARS-CoV-2-conserved-structured-45	26568–26687	UCCUUGAACAAUGGAACCUAGUAAUAGGUUUCCUAUUCCUUACAUGGAUUUGUCUUCUACAAUUUGCCUAUGCCAACAGGAAUAGGUUUUUGUAUAUAAUUA AGUUAAUUUUCCUCUGGC	..((((((((.((((((((....))))))))....(((......))).)))).....(((((...((((((.((....)).))))))..))))).....))))))................	−2.13	0.982
SARS-CoV-2-conserved-structured-46	1238–1357	AUUGUGGUGAAACUUCAUGGCAGACGGGCGAUUUUGUUAAAGCCACUUGCGAAUUUUGUGGCACUGAGAAUUUGACUAAAGAAGGUGCCACUACUUGUGGU UACUUACCCCAAAAUGCUG	.((((.((((....)))).))))...(((.((((((.....(((((...........)))))......................(((((((((.....)))).)))))....))))))))).	−0.75	0.981
SARS-CoV-2-conserved-structured-47	6451–6570	GAGUGUAAUGUGAAAACUACCGAAGUUGUAGGAGACAUUAUACUUAAACCAGCAAAUAAUAGUUUAAAAAUUACAGAAGAGGUUGGCCACACAGAUCUAAU GGCUGCUUAUGUAGACAAU	(((((((((((.....((((........))))...)))))))))))..........................(((((....((((.(((((............)))))))))).))))).....	−0.83	0.981
SARS-CoV-2-conserved-structured-48	29078–29197	AUGUAACACAAGCUUUCGGCAGACGUGGUCCAGAACAAACCCAAGGAAAUUUUGGGGACCAGGAACUAAUCAGACAAGGAACUGAUUACAAACAUUGGCCGC AAAUUGCACAAUUUGCCC	.................(((...(.((((((.(......)(((((.....))))))))))).)...(((((((.........))))))).........)))((((((((....))))))))..	−1.24	0.980
SARS-CoV-2-conserved-structured-49	23497–23616	CGUGCAGGCUGUUUAAUAGGGGCUGAACAUGUCAACAACUCAUAUGAGUGUGACAUACCCAUUGGUGCAGGUAUAUGCGCUAGUUAUCAGACUCAGACUAA UUCUCCUCGGCGGGCACGU	(((((..((((........(((.((((..((((((...((((....)))).))))))....((((((((((......)))))))))..)))).)))(((.....)))...))))..))))).	−2.42	0.979
SARS-CoV-2-conserved-structured-50	13686–13805	GAAGAAACAAUUUAUAAUUUACUUAAGGAUUGUCCAGCUGUUGCUAAACAUGACUUCUUUAAGUUUAGAAUAGACGGUGACAUGGUACCACAUAUAUCACGU CAACGUCUUACUAAAUAC	....................((((((((((.(((..(((....)))..))).....))))))))) ((((...(((((.((((.(((((........))))).)))).)))))..))))....	−3.1	0.978
SARS-CoV-2-conserved-structured-51	12610–12729	AUGGACAAUUCACCUAAUUUAGCAUGGCCUCUUAUUGUAACAGCUUUAAGGGCCAAUUCUGCUGUCAAAUUACAGAAUAAUGAGCUUAGUCCUGUUGCACUA CGACAGAUGUCUUGUGCU	..((((((((((.............(((((.((((..((....))..))))))))))((((((.((.......))))))))..)))).)).)))).....((((...(((....)))..)))).	−1.94	0.978
SARS-CoV-2-conserved-structured-52	19645–19764	UUGUAAAUAAGGGACACUUUGAUGGACAACAGGGUGAAGUACCAGUUUCUAUCAUUAAUAACACUGUUUACACAAAAGUUGAUGGUGUUGAUGUAGAAUUGU UUGAAAAUAAAACAACAU	...(((((((...(((...((((((((.(((.((.........)).))))))))))....((((((((((.((.......)).)))))))))).)))....)))))))................	−0.89	0.977
SARS-CoV-2-conserved-structured-53	11050–11169	CAGAGUACUCAAUGGUCUUUGUUCUUUUUUUUGUAUGAAAAUGCCUUUUUACCUUUUGCUAUGGGUAUUAUUGCUAUGUCUGCUUUUGCAAUGAUGUUUGUC AAACAUAAGCAUGCAUUU	(((((.(((....)))))))).........((((((.....((..(..((((((.........))))))..)..)) (((((.(((....))).((((....)))).)))))..))))))...	−0.33	0.977
SARS-CoV-2-conserved-structured-54	17445–17564	CAAUUACCUGCACCACGCACAUUGCUAACUAAGGGCACACUAGAACCAGAAUAUUUCAAUUCAGUGUGUAGACUUAUGAAAACUAUAGGUCCAGACAUGUUC CUCGGAACUUGUCGGCGU	.........((.((((((((..((((........))))...........((((.......)))).)))))) .((((((((......))))))))..((((.((((....)))).))))).))	−2.14	0.974
SARS-CoV-2-conserved-structured-55	24256–24375	GCUAUGCAAAUGGCUUAUAGGUUUAAUGGUAUUGGAGUUACACAGAAUGUUCUCUAUGAGAACCAAAAAUUGAUUGCCAACCAAUUUAAUAGUGCUAUUGGC AAAAUUCAAGACUCACUU	(((((....)))))...................(((((..........((((((...))))))......(((((((((((...(((....)))....))))))) ...))))))))....	−0.32	0.974
SARS-CoV-2-conserved-structured-56	17525–17644	AACUAUAGGUCCAGACAUGUUCCUCGGAACUUGUCGGCGUUGUCCUGCUGAAAUUGUUGACACUGUGAGUGCUUUGGUUUAUGAUAAUAAGCUUAAAGCACA UAAAGACAAAUCAGCUCA	......(((.((.((((.((((....)))).))))))......)))((((((..(((((........((.((((((((((((((....))))))).))))))))).))..))))).)))))...	−2.22	0.973

SARS-CoV-2-conserved-structured-57	27248–27367	AAUUAUUAUGAGGACUUUUAAAGUUUCCAUUUGG AAUCUUGAUUACAUCAUAAACCUCAUAAUUAAAA AUUUAUCUAAGUCACUAACUGAGAAUAAAUAUUC UCAAUUAGAUGAAGAGCA	....((((((((.......(((.((((....)))).)))(((...))) .....))))))))...........((...((((((. (((((((....))))))).)))).)))..))..	−3.72	0.972
SARS-CoV-2-conserved-structured-58	4574–4693	CACUUAACGAUCUAAAUGAAACUCUUGUUACAAU GCCACUUGGCUAUGUAACACAUGGCUUAAAUUUG GAAGAAGCUGCUCGGUAUAUGAGAUCUCUCAAAG UGCCAGCUACAGUUUCUG	..........((((((..((..(.((((((..(((....)))..)))))) ...)..))..))))))((((((((.(((((. ((((....))))..))))))))...)))))).	−3.73	0.971
SARS-CoV-2-conserved-structured-59	21325–21444	UCAUGCAUGCAAAUUACAUAUUUUGGAGGAAUAC AAAUCCAAUUCAGUUGUCUUCCUAUUCUUUAUUU GACAUGAGUAAAUUUCCCCUUAAAUUAAGGGGUA CUGCUGUUAUGUCUUUAA	.((((((((((((.....((...((((((.. ((............))..))))).)).....))))).)))) ((((.....(((((((...)))))))...))))..)))).......	−1.24	0.970
SARS-CoV-2-conserved-structured-60	7771–7890	CUUUACUUUGAUAAAGCUGGUCAAAAGACUUAUG AAAGACAUUCUCUCUCUCAUUUUGUUAACUUAGA CAACCUGAGAGCUAAUAACACUAAAGGUUCAUUG CCUAUUAAUGUUAUAGUU	.....((((.((((..((........))..)))).))))......... ((((((..((((((.......)))))..)))))) ...((((((. (((((((......)))).))).)))))))....	−1.93	0.967
SARS-CoV-2-conserved-structured-61	19765–19884	UACCUGUUAAUGUAGCAUUUGAGCUUUGGGCUAA GCGCAACAUUAAACCAGUACCAGAGGUGAAAAUA CUCAAUAAUUUGGGUGUGGACAUUGCUGCUAAUA CUGUGAUCUGGGACUACA	..(((.(((((((.((.(((.(((.....))))).)).))))))).((((((.(((..(((...((((((((....))))))))...)))..)))....))))).) ....))).....	−0.84	0.965
SARS-CoV-2-conserved-structured-62	6971–7090	UAAGUGUUUGCCUAGGUUCUUUAAUCUACUCAAC CGCUGCUUUAGGUGUUUUAAUGUCUAAUUUAGGC AUGCCUUCUUACUGUACUGGUUACAGAGAAGGCU AUUUGAACUCUACUAAUG	..((((.(((..((((((....))))))..))).)))).... ((((.((((.. (((((((...)))))))((((((((..(((((......))))))))))))) .....)))))))).......	−1.9	0.963
SARS-CoV-2-conserved-structured-63	28998–29117	AGGCCAAACUGUCACUAAGAAAUCUGCUGCUGAG GCUUCUAAGAAGCCUCGGCAAAAACGUACUGCCA CUAAAGCAUACAAUGUAACACAAGCUUUCGGCAG ACGUGGUCCAGAACAAAC	.((((...................... ((((((((((((...))))))))))))...((((.((((((... (((((..................))))).)))))))))))	−4.19	0.963
SARS-CoV-2-conserved-structured-64	5011–5130	GUUGUGGACAUGUCAAUGACAUAUGGACAACAGU UUGGUCCAACUUAUUUGGAUGGAGCUGAUGUUAC UAAAAUAAAACCUCAUAAUUCACAUGAAGGUAAA ACAUUUUAUGUUUUACCU	..((((((.(((((...)))))...((((.((((((.((((((......)))))) .)))))).)))).....................)))))).... (((((((((((...)))))))))))	−4.19	0.963
SARS-CoV-2-conserved-structured-65	4094–4213	CUUUCUUAAAGAAAGAUGCUCCAUAUAUAGUGGG UGAUGUUGUUCAAGAGGGUGUUUUAACUGCUGU GGUUAUACCUACUAAAAAGGCUGGUGGCACUACU GAAAUGCUAGCGAAAGCUU	..(((((.((.((...(((.((((.....))))))...)).)).))))) (((((((((.......((((..((((.. ((.....))..))))..))))))))))))(((.....))).	0.4	0.958
SARS-CoV-2-conserved-structured-66	1598–1717	GUCUUAAUGACAACCUUCUUGAAAUACUCCAAAA AGAGAAAGUCAACAUCAAUAUUGUUGGUGACUUU AAACUUAAUGAAGAGAUCGCCAUUAUUUUGGCAU CUUUUUCUGCUUCCACAA	(((......)))...((((((((.....(((.......))) (((((((.((. ((....)).)).)))))))....)))).))))((((. ((((.......))))))))................	−2	0.955
SARS-CoV-2-conserved-structured-67	11090–11209	AUGCCUUUUUACCUUUUGCUAUGGGUAUUAUUGC UAUGUCUGCUUUUGCAAUGAUGUUUGUCAAACAU AAGCAUGCAUUUCUCUGUUUGUUUUUGUUACCUU CUCUUGCCACUGUAGCUU	..................(((((((.(((((((((...(....) ....)))))))))........((((.((((((.(((.......))).))))).))))...............)))))))..	−1.63	0.955
SARS-CoV-2-conserved-structured-68	17965–18084	CUGAUAGAGACCUUUAUGACAAGUUGCAAUUUAC AAGUCUUGAAAUUCCACGUAGGAAUGUGGCAACU UUACAAGCUGAAAAUGUAACAGGACUCUUUAAAG AUUGUAGUAAGGUAAUCA	..(((....(((((.....(((((((((.(((...((. ((((((((((((((((....)))))).)).))).)))))).))).)))))) ......(((....)))))))...))))).))).	−0.96	0.952
SARS-CoV-2-conserved-structured-69	25815–25934	GAUGCCAACUAUUUUCUUUGCUGGCAUACUAAUU GUUACGACUAUUGUAUACCUUACAAUAGUGUAAC UUCUUCAAUUGUCAUUACUUCAGGUGAUGGCACA ACAAGUCCUAUUUCUGAA	.((((((.(...........).))))))........ (((((. ((((((((((......)))))))))))))).......... ((((((((((......)))))))))........................	−4.17	0.952
SARS-CoV-2-conserved-structured-70	14446–14565	AUGGUGUUCCAUUUGUAGUUUCAACUGGAUACCA CUUCAGAGAGCUAGGUGUUGUACAUAAUCAGGAU GUAAACUUACAUAGCUCUAGACUUAGUUUUAAGG AAUUACUUGUGUAUGCUG	..(((((..(((..((((((((..((((((.......))))))((((((((.(((...(((((.........)))))).....))).)))))))) ....))))))))..))).))))).	−1.41	0.948
SARS-CoV-2-conserved-structured-71	2118–2237	CACUGUUUAUGAAAAACUCAAACCCGUCCUUGAU UGGCUUGAAGAGAAGUUUAAGGAAGGUGUAGAG UUUCUUAGAGACGGUUGGGAAAUUGUUAAAUUUA UCUCAACCUGUGCUUGUGA	(((.((((.(((.((((((..(((..(((((((... ((((.......)))))))))))).)))...)))))).))).)))) (((((((((((((....))))..))))))))........))).	−2.75	0.948

(Continued)

TABLE 2. *(Continued)*

Name	Sequence interval	Sequence	Secondary structure	z-score	P value
SARS-CoV-2-conserved-structured-72	17205–17324	UAUUUGCCUAUAGAUAAAUGUAGUAGAAUUAUACCUGCACGUGCUCGUGUAGAGUGUUUUGAUAAAUUCAAAGUGAAUUCAACAUUAGAACAGUAUGUCUUU UGUACUGUAAAUGCAUUG	((((((....))))))((((((((.((((((((((((((((....))))))).))))(((((.....)))))...))))).))......((((((((.....))))))))...)))))).	−2.61	0.947
SARS-CoV-2-conserved-structured-73	7171–7290	UCUUUAGAAACUAUACAAAUUACCAUUUCAUCUUUUAAAUGGGAUUUAACUGCUUUUGGCUUAGUUGCAGAGUGGUUUUUGGCAUAUAUUCUUUUCACUAGG UUUUUCUAUGUACUUGGA	...((((((((((..(((((.((((((........))))))))))))..((((..(((...)))..)))).))))))))))................((((((.........)))))).	−0.53	0.945
SARS-CoV-2-conserved-structured-74	14766–14885	GCUCAGGAUGGUAAUGCUGCUAUCAGCGAUUAUGACUACUAUCGUUAUAAUCUACCAACAAUGUGUGAUAUCAGACAACUACUAUUUGUAGUUGAAGUUGUU GAUAAGUACUUUGAUUGU	(((...(((((((....))))))))((((((((.......))))))))...((((((((.((....)).)) ((((((.....))))))...)))))))..................	−2.13	0.943
SARS-CoV-2-conserved-structured-75	11730–11849	UAUGAAUUCACAGGGACUACUCCCACCCAAGAAUAGCAUAGAUGCCUUCAAACUCAACAUUAAAUUGUUGGGUGUUGGUGGCAAACCUUGUAUCAAAGUAG CCACUGUACAGUCUAAAAU	...((....((((((.(((((......((((...........((((.((((((((((((......))))))))).)))).))))...))))......))))))).)))).....)).......	−1.59	0.942
SARS-CoV-2-conserved-structured-76	29238–29357	GGAAGUCACACCUUCGGGAACGUGGUUGACCUACACAGGUGCCAUCAAAUUGGAUGACAAAGAUCCAAAUUUCAAAGAUCAAGUCAUUUUGCUGAAUAAGCA UAUUGACGCAUACAAAAC	....((((.(((..((....)).)))))))........((((.......((((((........)))))) ..............((((...((((.....))))...)))))))).......	−1.58	0.941
SARS-CoV-2-conserved-structured-77	1558–1677	GGUUGUAACCAUACAGGUGUUGUUGGAGAAGGUUCCGAAGGUCUUAAUGACAACCUUCUUGAAAUACUCCAAAAAGAGAAAGUCAACAUCAAUAUUGUUGG UGACUUUAAACUUAAUGAA	(((....))).....((.((..((.((((((((.......(((.....)))))))))))).))..)).))... (((..((((((((.((.((....)).)).)))))))...)))......	−1.18	0.941
SARS-CoV-2-conserved-structured-78	4214–4333	UGAGAAAAGUGCCAACAGACAAUUAUAUAACCACUUACCCGGGUCAGGGUUUAAAUGGUUACACUGUAGAGGAGGCAAAGACAGUGCUUAAAAAGUGUAAAA GUGCCUUUUACAUUCUAC	..((((.((((.................((((((...((((......)))).....)))))))))) (((((..(((((...(((.(.........).))).....))))))))).))))..	−0.39	0.940
SARS-CoV-2-conserved-structured-79	9410–9529	CUGGUGGUAUUGUAGCUAUCGUAGUAACAUGCCUUGCCUACUAUUUUAUGAGGUUUAGAAGAGCUUUUGGUGAAUACAGUCAUGUAGUUGCCUUUAAUACUU UACUAUUCCUUAUGUCAU	..((((((((.((.((((...)))).))))))))..)))........(((((((..(((.((((..((.((..(.((((....)))).)..))...))..)))).)))..))))))))....	−1.74	0.940
SARS-CoV-2-conserved-structured-80	13806–13925	ACAAUGGCAGACCUCGUCUAUGCUUUAAGGCAUUUUGAUGAAGGUAAUUGUGACACAUUAAAAGAAAUACUUGUCACAUACAAUUGUUGUGAUGAUGAUUAU UUCAAUAAAAAGGACUGG	(((((......((((((((.(((((....)))))...)))).)))).))))).............((((((.(..((((((((.....).)))).)))..).))))))..................	−1.73	0.940
SARS-CoV-2-conserved-structured-81	23537–23656	CAUAUGAGUGUGACAUACCCAUUGGUGCAGGUAUAUGCGCUAGUUAUCAGACUCAGACUAAUUCUCCUCGGCGGGCACGUAGUGUAGCUAGUCAAUCCAUCA UUGCCUACACUAUGUCAC	....(((((.(((........(((((((((......))))))))))..))).)))))...........((.....)) (.(((((((((((.((((..........)))).))))))))))))..	−0.51	0.940
SARS-CoV-2-conserved-structured-82	9210–9329	GUACUGUAGGCACGGCACUUGUGAAAGAUCAGAAGCUGGUGUUUGUGUAUCUACUAGUGGUAGAUGGGUACUUAACAAUGAUUAUUACAGAUCUUUACCAG GAGUUUUCUGUGGUGUAGA	(((((.(((((((.(..(((.(((....))).))).).)))))))))..(((((((......))))))))))))...........((((((((.((((....))))...))))))))))....	−0.31	0.939
SARS-CoV-2-conserved-structured-83	6091–6210	CAGUUAACUGGUUAUAAGAAACCUGCUUCAAGAGAGCUUAAAGUUACAUUUUUCCCUGACUUAAAUGGUGAUGUGGUGGCUAUUGAUUAUAAACACUACACA CCCUCUUUUAAGAAAGGA	..((.((((((((.......)))).((((.....))))....)))))).......(((..((((((.((((.(((((((..(((......)))..))))))))))).....))))))..))).	−2.57	0.939
SARS-CoV-2-conserved-structured-84	29118–29237	CCAAGGAAAUUUUGGGGACCAGGAACUAAUCAGACAAGGAACUGAUUACAAACAUUGGCCGCAAAUUGCACAAUUUGCCCCCAGCGCUUCAGCGUUCUUCGG AAUGUCGCGCAUUGGCAU	((((.........(((((.((((....(((((((.........))))))).......))))..(((((((.....))))))))))))((((.....((((((.....)))))).)))).))))...	−3.14	0.938
SARS-CoV-2-conserved-structured-85	28958–29077	AGCUUGAGAGCAAAAUGUCUGGUAAAGGCCAACAACAACAAGGCCAAACUGUCACUAAGAAAUCUGCUGCUGAGGCUUCUAAGAAGCCUCGGCAAAAACGUA CUGCCACUAAAGCAUACA	.((((.((.(((..((((..((((..((((............))))....((........)).....))))((((((((((((...)))))))))))).....))))..)))..)).)))))......	−2.78	0.935

SARS-CoV-2-conserved-structured-86	29278–29397	GCCAUCAAAUUGGAUGACAAAGAUCCAAAUUUCA AAGAUCAAGUCAUUUUGCUGAAUAAGCAUAUUGA CGCAUACAAAACAUUCCCACCAACAGAGCCUAAA AAGGACAAAAAGAAGAAG	.........((((((.......)))))).((((.........((((... ((((.....))))...))))........................ ((((....))).).....))))...	−1.58	0.935
SARS-CoV-2-conserved-structured-87	14046–14165	GGUGUACUGACAUUAGAUAAUCAAGAUCUCAAUG GUAACUGGUAUGAUUUCGGUGAUUUCAUACAAAC CACGCCAGGUAGUGGAGUUCCUGUUGUAGAUUCU UAUUAUUCAUUGUUAAUG	.......(((((.(..(((((.((((((((((((.((((.((((((.. ((...))..))))))... ((((........))))))).)))))).))).)))))))))..).)))))...	−2.05	0.935
SARS-CoV-2-conserved-structured-88	9250–9369	GUUUGUGUAUCUACUAGUGGUAGAUGGGUACUUA ACAAUGAUUAUUACAGAUCUUUACCAGGAGUUUU CUGUGGUGUAGAUGCUGUAAAUUUACUUACUAAU AUGUUUACACCACUAAUU	(((...((((((((((....))).)))))))..)))..........(((((. ((((....))))...)))))((((((((...((((......)))) ......)))))))).......	−0.53	0.934
SARS-CoV-2-conserved-structured-89	26415–26527	GUUUACUCUCGUGUUAAAAAUCUGAAUUCUUCUA GAGUUCCUGAUCUUCUGGUCUAAACGAACUAAAU AUUAUAUUAGUUUUUCUGUUUGGAACUUUAAUUU UAGCCAUGGCA	........((((((((((....((.....))((((((((.((((....)))). (((((((((((........)))))) ....)))))))))))).)))))).)))))..	−1.94	0.934
SARS-CoV-2-conserved-structured-90	14846–14965	ACUAUUUGUAGUUGAAGUUGUUGAUAAGUACUUU GAUUGUUACGAUGGUGGCUGUAUUAAUGCUAACC AAGUCAUCGUCAACAACCUAGACAAAUCAGCUGG UUUUCCAUUUAAUAAAUG	..........((((((...((((((..((.((..(.(((((.. (((((((((...(((....))) ...)))))))))))))))..))))..))))))(((....)))))))))	−0.79	0.934
SARS-CoV-2-conserved-structured-91	4971–5090	GGUGUUUACAACAGUAGACAACAUUAACCUCCAC ACGCAAGUUGUGGACAUGUCAAUGACAUAUGGAC AACAGUUUGGUCCAACUUAUUUGGAUGGAGCUGA UGUUACUAAAAUAAAACC	(((((((((....))))))(((((...((((((((....).)))))). (((((...)))))..))....((((((. ((((((.....)))))).))))))))))...........)))	−3.72	0.932
SARS-CoV-2-conserved-structured-92	9610–9729	CUUACUAAUGAUGUUUCUUUUUUAGCACAUAUUC AGUGGAUGGUUAUGUUCACACCUUUAGUACCUUU CUGGAUAACAAUUGCUUAUAUCAUUUGUAUUUCC ACAAAGCAUUUCUAUUGG	.......(((.((((........)))))))..... ((((((.......)))))).((..(((...((((.((((..((((. ((........)).))))...)))).)))).....)))..))	−0.85	0.932
SARS-CoV-2-conserved-structured-93	5691–5810	ACCACCUGCUCAGUAUGAACUUAAGCAUGGUACA UUUACUUGUGCUAGUGAGUACACUGGUAAUUACC AGUGUGGUCACUAUAAACAUAUAACUUCUAAAGA AACUUUGUAUUGCAUAGA	...(((((((.(((....)))..)))).)))........(((((((((((. (((((((((....))))))))).))))))......((((((.(((....))) ...)))))).))))))).	−3.18	0.931
SARS-CoV-2-conserved-structured-94	25016–25135	UAGAUAAAUAUUUUAAGAAUCAUACAUCACCAGA UGUUGAUUUAGGUGACAUCUCUGGCAUUAAUGCU UCAGUUGUAAACAUUCAAAAAGAAAUUGACCGCC UCAAUGAGGUUGCCAAGA	.((((...........((((((.(((((....)))))))))))).(....)))). (((((........((((((..................)))))).. (((((...)))))))))...	−1.16	0.926
SARS-CoV-2-conserved-structured-95	6411–6530	AGAAGUAGUGGAAAAUCCUACCAUACAGAAAGAC GUUCUUGAGUGUAAUGUGAAAACUACCGAAGUUG UAGGAGACAUUAUACUUAAACCAGCAAAUAAUAG UUUAAAAAUUACAGAAGA	....(((((((.....))................(((. ((((((((((((.....((((........))))...)))))))))))) ...)))..................)))))......	−1.1	0.922

SARS-CoV-2-conserved-structured-96	26608–26727	UACAUGGAUUUGUCUUCUACAAUUUGCCUAUGCC
		AACAGGAAUAGGUUUUUGUAUAUAAUUAAGUUAA
		UUUUCCUCUGGCUGUUAUGGCCAGUAACUUUAGC UUGUUUUGUGCUUGCUGC
			..((..((.........(((((...((((((.((....)).))))))..)))))
			......((((((((........((((((.....))))))
			.....)))))))).))..))........
				−1.64	0.922

SARS-CoV-2-conserved-structured-97	11010–11129	GUUGUUACUCACAAUUUUGACUUCACUUUUAGUU UUAGUCCAGAGUACUCAAUGGUCUUUGUUCUUUU UUUUGUAUGAAAAUGCCUUUUUACCUUUUGCUAU GGGUAUUAUUGCUAUGUC	..((.(((((...(((..((((.........))))..)))...)))))..)) (((((..................((((....))))...... (((((.........)))))......)))))...	0.12	0.922
SARS-CoV-2-conserved-structured-98	10850–10969	CUUCAUUAAAAGAAUUACUGCAAAAUGGUAUGAA UGGACGUACCAUAUUGGGUAGUGCUUUAUUAGAA GAUGAAUUUACACCUUUUGAUGUUGUUAGACAAU GCUCAGGUGUUACUUUCC	((((....((((.((((((.(((. ((((((((.......)))))))).)))))))).))))....)))).. (((...((((((...(.(((((......)))))..))))))....))).	−2.67	0.920

SARS-CoV-2-conserved-structured-99	14886–15005	UACGAUGGUGGCUGUAUUAAUGCUAACCAAGUCA
		UCGUCAACAACCUAGACAAAUCAGCUGGUUUUCC
		AUUUAAUAAAUGGGGUAAGGCUAGACUUUAUUAU GAUUCAAUGAGUUAUGAG
			...(((((((((...(((.....)))...)))))))))
			................((((((....(((((((......)))))))
			...)))).))......((((((((((...)))))))))).
				−1.33	0.919

(Continued)

TABLE 2. *(Continued)*

Name	Sequence interval	Sequence	Secondary structure	z-score	P value
SARS-CoV-2-conserved-structured-100	25216–25335	GGUUUUAUAGCUGGCUUGAUUGCCAUAGUAAUGG UGACAAUUAUGCUUUGCUGUAUGACCAGUUGCUG UAGUUGUCUCAAGGGCUGUUGUUCUUGUGGAUCC UGCUGCAAAUUUGAUGAA	((((.((((((.(((.(((((((((....)))) ...)))))).)))..)))))).))))..((((.((((.. ((((((((((....))))))).)))).)))).))))..........	−3.78	0.916
SARS-CoV-2-conserved-structured-101	11690–11809	GUGUUUAUGAUUACUUAGUUUCUACACAGGAGUU UAGAUAUAUGAAUUCACAGGGACUACUCCCACCC AAGAAUAGCAUAGAUGCCUUCAAACUCAACAUUA AAUUGUUGGGUGUUGGUG	((((((((....(((.............(((((((......)))))))... ((((....))))....)))....))))))))).. (((((((((((......)))))))).))))..	−1.17	0.913
SARS-CoV-2-conserved-structured-102	5251–5370	GGUUUAACUUCUAUUAAAUGGGCAGAUAACAACU GUUAUCUUGCCACUGCAUUGUUAACACUCCAACA AAUAGAGUUGAAGUUUAAUCCACCUGCUCUACAA GAUGCUUAUUACAGAGCA	((...((((((.(((.....((((((((((....)))))).)))).(((.. (((((........))))).)))))).))))))....))....(((((.. ((........))..))))).	−2.28	0.912
SARS-CoV-2-conserved-structured-103	1198–1317	UGCAACCAAAUGUGCCUUUCAACUCUCAUGAAGU GUGAUCAUUGUGGUGAAACUUCAUGGCAGACGG GCGAUUUUGUUAAAGCCACUUGCGAAUUUUGUGG CACUGAGAAUUUGACUAAA	......(((((((((...(((....((((((((.(. (((.....))).).))))))))((((... (((.............)))..))))......))).)))))......)))))	−0.42	0.910
SARS-CoV-2-conserved-structured-104	28758–28877	UCAAGGAACAACAUUGCCAAAAGGCUUCUACGCA GAAGGGAGCAGAGGCGGCAGUCAAGCCUCUUCU CGUUCCUCAUCACGUAGUCGCAACAGUUCAAGAA AUUCAACUCCAGGCAGCAG	.............(((((.....((..(((((.....((((((((((. (((.......))).))))).))))))......)))))..))...((((. ((....)).))))...)))))...	−3.57	0.910
SARS-CoV-2-conserved-structured-105	17765–17884	CUUUAUUUCACCUUAUAAUUCACAGAAUGCUGUA GCCUCAAAGAUUUUGGGACUACCAACUCAAACUG UUGAUUCAUCACAGGGCUCAGAAUAUGACUAUGUCAUAUUCACUCAAACCAC	......................((((....))))((((.....((.. ((((.....)))).))...((((.((.....))))))))))... (((((((((...)))))))))...........	−2.03	0.903
SARS-CoV-2-conserved-structured-106	22101–22218	AGGAAAACAGGGUAAUUUCAAAAAUCUUAGGGAA UUUGUGUUUAAGAAUAUUGAUGGUUAUUUUAAAA UAUAUUCUAAGCACACGCCUAUUAAUUUAGUGCG UGAUCUCCCUCAGGGU	........................(((((((((....(((((((. ((((((...................)))))))))))) (((((((.......))).))))....))))).)))))	−1.7	0.900

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 nt. Sequence intervals are relative to the reference genome NC_045512.2.

TABLE 3. SARS-CoV-2-conserved-unstructured

Name	Interval	Average unpaired probability	Minimum unpaired proability	Sequence
SARS-CoV-2-conserved-unstructured-1	29074-29087	0.891	0.764	AUACAAUGUAACAC
SARS-CoV-2-conserved-unstructured-2	8078-8094	0.825	0.753	CCAAUGGAAAAACUCAA
SARS-CoV-2-conserved-unstructured-3	1359-1374	0.837	0.717	UUGUUAAAAUUUAUUG
SARS-CoV-2-conserved-unstructured-4	21626-21643	0.857	0.713	ACUCAAUUACCCCCUGCA
SARS-CoV-2-conserved-unstructured-5	1420-1436	0.797	0.697	CGAAUACCAUAAUGAAU
SARS-CoV-2-conserved-unstructured-6	18471-18484	0.780	0.695	UCAAUUUAAACACC
SARS-CoV-2-conserved-unstructured-7	11910-11923	0.767	0.683	AAUCAUCAUCUAAA
SARS-CoV-2-conserved-unstructured-8	23960-23981	0.787	0.678	UUUAAUUUUUCACAAAUAUUAC
SARS-CoV-2-conserved-unstructured-9	13990-14003	0.796	0.662	CAAGCUUUGUUAAA
SARS-CoV-2-conserved-unstructured-10	10009-10035	0.760	0.657	UCUUUACCAACCACCACAAACCUCUAU
SARS-CoV-2-conserved-unstructured-11	23700-23718	0.823	0.655	CCAUACCCACAAAUUUUAC
SARS-CoV-2-conserved-unstructured-12	18918-18934	0.832	0.654	UAUUGAAUAUCCUAUAA
SARS-CoV-2-conserved-unstructured-13	27385-27402	0.810	0.654	UAAACGAACAUGAAAAUU
SARS-CoV-2-conserved-unstructured-14	5773-5789	0.808	0.654	UAAACAUAUAACUUCUA
SARS-CoV-2-conserved-unstructured-15	23910-23932	0.838	0.653	CACAAGUCAAACAAAUUUACAAA
SARS-CoV-2-conserved-unstructured-16	17762-17785	0.767	0.650	CUGUCUUUAUUUCACCUUAUAAUU
SARS-CoV-2-conserved-unstructured-17	25569-25582	0.826	0.649	UUCCAAAAUCAUAA
SARS-CoV-2-conserved-unstructured-18	19569-19588	0.754	0.648	UUACAAACAAUUUGAUACUU
SARS-CoV-2-conserved-unstructured-19	22552-22565	0.773	0.647	UAAUAUUACAAACU
SARS-CoV-2-conserved-unstructured-20	25417-25437	0.747	0.640	ACAAUUGGAACUGUAACUUUG
SARS-CoV-2-conserved-unstructured-21	12195-12210	0.746	0.634	UUAAAAAGUUGAAGAA
SARS-CoV-2-conserved-unstructured-22	6757-6783	0.790	0.633	UUGUACUAAUUAUAUGCCUUAUUUCUU
SARS-CoV-2-conserved-unstructured-23	15236-15257	0.747	0.630	ACAACAUGUUAAAAACUGUUUA
SARS-CoV-2-conserved-unstructured-24	6225-6238	0.734	0.628	AUAAACCUAUUGUU
SARS-CoV-2-conserved-unstructured-25	9578-9598	0.826	0.628	UAUUCUGUUAUUUACUUGUAC
SARS-CoV-2-conserved-unstructured-26	21649-21662	0.821	0.627	UAAUUCUUUCACAC

Continued

TABLE 3. *(Continued)*

Name	Interval	Average unpaired probability	Minimum unpaired proability	Sequence
SARS-CoV-2-conserved-unstructured-27	23985-23998	0.800	0.625	AUCCAUCAAAACCA
SARS-CoV-2-conserved-unstructured-28	7161-7174	0.774	0.625	ACACCUAUCCUUCU
SARS-CoV-2-conserved-unstructured-29	6010-6029	0.739	0.625	ACCAAACCAACCAUAUCCAA
SARS-CoV-2-conserved-unstructured-30	6515-6529	0.811	0.624	UUAAAAAUUACAGAA
SARS-CoV-2-conserved-unstructured-31	18219-18232	0.750	0.624	UAAAAUGAAUUAUC
SARS-CoV-2-conserved-unstructured-32	11659-11681	0.819	0.623	UUUACUCAACCGCUACUUUAGAC
SARS-CoV-2-conserved-unstructured-33	24778-24797	0.771	0.622	AAAGAACUUCACAACUGCUC
SARS-CoV-2-conserved-unstructured-34	21669-21683	0.789	0.621	UUUAUUACCCUGACA
SARS-CoV-2-conserved-unstructured-35	6105-6122	0.777	0.619	AUAAGAAACCUGCUUCAA
SARS-CoV-2-conserved-unstructured-36	28436-28452	0.808	0.619	GCUCUCACUCAACAUGG
SARS-CoV-2-conserved-unstructured-37	16361-16375	0.773	0.619	UCUUGUCUGUUAAUC
SARS-CoV-2-conserved-unstructured-38	28148-28162	0.734	0.617	UUUUACAAUUAAUUG
SARS-CoV-2-conserved-unstructured-39	24165-24178	0.745	0.616	AAAUGAUUGCUCAA
SARS-CoV-2-conserved-unstructured-40	26429-26447	0.741	0.615	UUAAAAAUCUGAAUUCUUC
SARS-CoV-2-conserved-unstructured-41	6072-6086	0.727	0.615	AAUUUGCUGAUGAUU
SARS-CoV-2-conserved-unstructured-42	1304-1319	0.822	0.615	GAGAAUUUGACUAAAG
SARS-CoV-2-conserved-unstructured-43	1918-1933	0.752	0.614	UCUUGAAACUGCUCAA
SARS-CoV-2-conserved-unstructured-44	27361-27375	0.741	0.612	GAAGAGCAACCAAUG
SARS-CoV-2-conserved-unstructured-45	28853-28866	0.770	0.612	UCAAGAAAUUCAAC
SARS-CoV-2-conserved-unstructured-46	14899-14913	0.742	0.610	UGUAUUAAUGCUAAC
SARS-CoV-2-conserved-unstructured-47	19260-19273	0.761	0.610	UAACCUUAACUUGC
SARS-CoV-2-conserved-unstructured-48	11724-11740	0.730	0.608	UUAGAUAUAUGAAUUCA
SARS-CoV-2-conserved-unstructured-49	29008-29023	0.780	0.608	UGUCACUAAGAAAUCU
SARS-CoV-2-conserved-unstructured-50	11537-11554	0.771	0.607	AUUGUUUUUAUGUGUGUU
SARS-CoV-2-conserved-unstructured-51	11628-11645	0.850	0.607	AUUUUUGUACUUGUUACU
SARS-CoV-2-conserved-unstructured-52	18681-18694	0.747	0.606	CACAUGCUUUUCCA
SARS-CoV-2-conserved-unstructured-53	7366-7384	0.724	0.606	AAUAAUUAAUCUUGUACAA

Continued

TABLE 3. (*Continued*)

Name	Interval	Average unpaired probability	Minimum unpaired proability	Sequence
SARS-CoV-2-conserved-unstructured-54	1031-1047	0.728	0.605	GAAAUUAAAUUGGCAAA
SARS-CoV-2-conserved-unstructured-55	14367-14380	0.714	0.605	UUGUGCAAACUUUA
SARS-CoV-2-conserved-unstructured-56	3797-3816	0.742	0.603	UUUGAUAAAAAUCUCUAUGA
SARS-CoV-2-conserved-unstructured-57	22281-22296	0.742	0.601	CUUUACUUGCUUUACA
SARS-CoV-2-conserved-unstructured-58	16038-16053	0.780	0.600	UUACCCACUUACUAAA

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.

were computed with CONTRAfold 2.0, and these were then used to calculate the expected number of true positive, true negative, false positive, and false negative base pairs. These computations were carried out using the Arnie package (https://github.com/DasLab/arnie).

RNAz (Gruber et al. 2010) structures were predicted in windows of the SARS-CoV-2 genome using the SARSr-MSA-1 alignment. We used rnazWindow.pl to compile alignment windows across SARSr-MSA-1 with at least four sequences in each window, using a window size of 120 nt sliding by 40 nt, and using default settings otherwise. RNAz hits were computed at the $P > 0.5$ threshold for the forward strand with z-scores computed without a shuffled sequence background for efficiency, using the –no-shuffle flag. The resulting RNAz structured windows were then clustered with rnazCluster.pl, filtered with rnazFilter.pl at a $P > 0.9$ threshold, and sorted with rnazSort.pl.

TABLE 4. Sequences and homology-modeled secondary structures for key structured genome regions of SARS-CoV-2

Name	Sequence	Secondary structure dot-bracket
Extended 5′ UTR	AUUAAAGGUUUAUACCUUCCCAGGUAACAAACCA ACCAACUUUCGAUCUCUUGUAGAUCUGUUCUCUA AACGAACUUUAAAAUCUGUGUGGCUGUCACUCGG CUGCAUGCUUAGUGCACUCACGCAGUAUAAUUAA UAACUAAUUACUGUCGUUGACAGGACACGAGUAA CUCGUCUAUCUUCUGCAGGCUGCUUACGGUUUCG UCCGUGUUGCAGCCGAUCAUCAGCACAUCUAGGU UUCGUCCGGGUGUGACCGAAAGGUAAGAUGGAG AGCCUUGUCCCUGGUUUCAACGAGAAAACACACG UCCAACUCAGUUUGCCUGUUUUACAGGUUCGCGA CGUGCUCGUACGUGGCUUUGGAGACUCCGUGGA GGAGGUCUUAUCAGAGGCACGUCAACAUCUUAAA GAUGGCACUUGUGGCUUAGUAGAAGUUGAAAAA GGCGUUUUGCC	......(((((.(((((....)))))..)))))......... ..(((((.....))))). ((((.......))))........(((((((.((.((((.(((... ..))).)))))).)))))))(.. (((((.....)))))...((((((((((..(((((...(((.((((((((((.((((((.(((((... ...)))))..)))))).....)))(((((((.((......)))))))) (((....))))))))))))).))))).))))...))))))).......((((((...... .((.. ((((((...))))))..))))))))(... .(((((. ((((((((((((.....)))).)))))..))))).)))))... ((((((...))))))...... (((((... ..)))))......(((.....)))
Frameshifting stimulation element	GUUUUUAAACGGGUUUGCGGUGUAAGUGCAGCC CGUCUUACACCGUGCGGCACAGGCACUAGUACUG AUGUCGUAUACAGGGCUUUUG	.... (((((((((((...[[[[[[[)))))))))))((((((((...... ...))).)))))...]].]]]]]....
3′ UTR	GACCACACAAGGCAGAUGGGCUAUAUAAACGUUU UCGCUUUUCCGUUUACGAUAUAUAGUCUACUCUU GUGCAGAAUGAAUUCUCGUAACUACAUAGCACAA GUAGAUGUAGUUAACUUUAAUCUCACAUAGCAAU CUUUAAUCAGUGUGUAACAUUAGGGAGGACUUGA AAGAGCCACCACAUUUUCACCGAGGCCACGCGGA GUACGAUCGAGUGUACAGUGAACAAUGCUAGGG AGAGCUGCCUAUAUGGAAGAGCCCUAAUGUGUAA AAUUAAUUUUAGUAGUGCUAUCCCCAUGUGAUUU UAAUAGCUUCUUAGGAGAAUGACAAAAAAAAAAA AAAAAAAAAAAAAAAAAAAAAAA	....(((((((((((..(((...((......))...))).))))))))))).. (((((((...... .{{{{{{.[[[[[[[[.)))))))...]]]]]]]]... ((((..((((((((((.. ((..(((.(((.....((((((((.((.(((... .)))))....(((((((((....((.. ((((.....))..))...))...))))).)))).((((........))))...... ....)))))))))).....))).)))..))...)))))... ...)))))..)))) ....}}}}}}... ....

We ran alifoldz (Washietl and Hofacker 2004) on the same genome windows used with RNAz above, again using SARSr-MSA-1. The alifoldz z-score computations were calculated for the forward strand only with alifoldz.pl. We additionally calculated alifoldz z-scores for alignment windows that were shuffled with shuffle-aln.pl to assess background z-scores, determining that 1% of shuffled alignment *z*-scores were less than −2.69.

We computed alignment powers with R-scape (Rivas et al. 2017) to assess the potential for using SARSr-MSA-1 for covariation analysis. We generated Stockholm alignment files with biopython (Cock et al. 2009) for windows of 120 nt each sliding by 40 nt. For each window, we ran R-scape with the –fold flag to predict new structures, obtaining estimates for the power of each base pair in the predicted structure (here, power is the expected sensitivity for detecting base pairs given the number of substitutions in the alignment at that base pair). We then averaged across base pairs in each structure to obtain the alignment power as described previously (Rivas et al. 2020), noting that all windows' alignment powers fell below the 0.10 threshold used to distinguish low-power from high-power alignments.

Analysis of unstructured elements

To obtain probabilities that each genome position in SARS-CoV-2 was unpaired, we computed base-pairing probability matrices with CONTRAfold 2.0 (Do et al. 2006) in windows of 120 nt sliding by 40 nt, and for each genome position, we summed the probabilities of pairing with all potential partners. We then averaged these nucleotide pairing probabilities across all windows that nucleotide was present in. Additionally, RNAplfold (Bernhart et al. 2006) was run with window size 120 nt, producing another set of unpaired probabilities for each position in the genome.

DATA DEPOSITION

Code used for the conservation, structured, and unstructured analyses above can be found at the GitHub repository: https://github.com/DasLab/SARSCoV2_Secstruct_Cons. The repository additionally includes alignment files, Rfam families and covariance models, and output from the RNAz, R-scape, alifoldz and RNAplfold analyses.

SUPPLEMENTAL MATERIAL

Supplemental material is available for this article.

ACKNOWLEDGMENTS

The authors acknowledge support from the National Science Foundation Graduate Research Fellowship Program (R.R., H.W.S.), a Stanford Graduate Fellowship (I.N.Z.), Stanford ChEM-H Postdocs at the Interface Seed Grant Award (R.J.H.), NIH/NIDDK 5T32DK007056-44 (E.A.P.), Stanford ChEM-H Physician Scientist Research Fellowship (E.A.P.), NIH/NIAID 1R01AI13219101 (J.S.G.), USAMRAA DOD W81XWH1810647 (J.S.G.), Harrington Scholar Innovator Grant (J.S.G.), Open Philanthropy (J.S.G.), and National Institutes of Health grant MIRA R35 GM122579 (R.D.). We would like to thank Andrew M. Watkins for updates to the RiboDraw software for rendering secondary structures and Dr. Paul Gardner for providing rapid reviews of the initial preprint of this manuscript on bioRxiv.

Received April 28, 2020; accepted May 11, 2020.

REFERENCES

Andrews RJ, Roche J, Moss WN. 2018. ScanFold: an approach for genome-wide discovery of local RNA structural elements-applications to Zika virus and HIV. *PeerJ* **6:** e6136. doi:10.7717/peerj.6136

Andrews RJ, Peterson JM, Haniff HS, Chen J, Williams C, Grefe M, Disney MD, Moss WN. 2020. An *in silico* map of the SARS-CoV-2 RNA Structurome. bioRxiv doi:10.1101/2020.04.17.045161

Bennett CF, Krainer AR, Cleveland DW. 2019. Antisense oligonucleotide therapies for neurodegenerative diseases. *Annu Rev Neurosci* **42:** 385–406. doi:10.1146/annurev-neuro-070918-050501

Bernhart SH, Hofacker IL, Stadler PF. 2006. Local RNA base pairing probabilities in large sequences. *Bioinformatics* **22:** 614–615. doi:10.1093/bioinformatics/btk014

Bustin SA, Nolan T. 2004. Pitfalls of quantitative real-time reverse-transcription polymerase chain reaction. *J Biomol Tech* **15:** 155–166. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2291693/

Ceraolo C, Giorgi FM. 2020. Genomic variance of the 2019-nCoV coronavirus. *J Med Virol* **92:** 522–528. doi:10.1002/jmv.25700

Chen S-C, Olsthoorn RCL. 2010. Group-specific structural features of the 5′-proximal sequences of coronavirus genomic RNAs. *Virology* **401:** 29–41. doi:10.1016/j.virol.2010.02.007

Cock PJ, Antao T, Chang JT, Chapman BA, Cox CJ, Dalke A, Friedberg I, Hamelryck T, Kauff F, Wilczynski B, et al. 2009. Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics* **25:** 1422–1423. doi:10.1093/bioinformatics/btp163

Connelly CM, Moon MH, Schneekloth JS Jr. 2016. The emerging role of RNA as a therapeutic target for small molecules. *Cell Chem Biol* **23:** 1077–1090. doi:10.1016/j.chembiol.2016.05.021

Do CB, Woods DA, Batzoglou S. 2006. CONTRAfold: RNA secondary structure prediction without physics-based models. *Bioinformatics* **22:** e90–e98. doi:10.1093/bioinformatics/btl246

Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* **32:** 1792–1797. doi:10.1093/nar/gkh340

Elbe S, Buckland-Merrett G. 2017. Data, disease and diplomacy: GISAID's innovative contribution to global health. *Glob Chall* **1:** 33–46. doi:10.1002/gch2.1018

Fehr AR, Perlman S. 2015. Coronaviruses: an overview of their replication and pathogenesis. *Methods Mol Biol* **1282:** 1–23. doi:10.1007/978-1-4939-2438-7_1

Goebel SJ, Hsue B, Dombrowski TF, Masters PS. 2004. Characterization of the RNA components of a putative molecular switch in the 3′ untranslated region of the murine coronavirus genome. *J Virol* **78:** 669–682. doi:10.1128/JVI.78.2.669-682.2004

Goebel SJ, Miller TB, Bennett CJ, Bernard KA, Masters PS. 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:** 1274–1287. doi:10.1128/JVI.00803-06

Gruber AR, Findeiß S, Washietl S, Hofacker IL, Stadler PF. 2010. RNAz 2.0: improved noncoding RNA detection. *Pac Symp Biocomput* 69–79. doi:10.1142/9789814295291_0009

Guan B-J, Su Y-P, Wu H-Y, Brian DA. 2012. Genetic evidence of a long-range RNA-RNA interaction between the genomic 5′ untranslated region and the nonstructural protein 1 coding region in murine and bovine coronaviruses. *J Virol* **86:** 4631–4643. doi:10.1128/JVI.06265-11

Hamada M, Sato K, Asai K. 2010. Prediction of RNA secondary structure by maximizing pseudo-expected accuracy. *BMC Bioinformatics* **11:** 586. doi:10.1186/1471-2105-11-586

Herold J, Siddell SG. 1993. An 'elaborated' pseudoknot is required for high frequency frameshifting during translation of HCV 229E polymerase mRNA. *Nucleic Acids Res* **21:** 5838–5842. doi:10.1093/nar/21.25.5838

Jonassen CM, Jonassen TO, Grinde B. 1998. A common RNA motif in the 3′ end of the genomes of astroviruses, avian infectious bronchitis virus and an equine rhinovirus. *J Gen Virol* **79:** 715–718. doi:10.1099/0022-1317-79-4-715

Kalvari I, Argasinska J, Quinones-Olvera N, Nawrocki EP, Rivas E, Eddy SR, Bateman A, Finn RD, Petrov AI. 2018. Rfam 13.0: shifting to a genome-centric resource for non-coding RNA families. *Nucleic Acids Res* **46:** D335–D342. doi:10.1093/nar/gkx1038

Kappel K, Zhang K, Su Z, Kladwang W, Li S, Pintilie G, Topkar VV, Rangan R, Zheludev IN, Watkins AM, et al. 2019. Ribosolve: rapid determination of three-dimensional RNA-only structures. bioRxiv doi:10.1101/717801

Katoh K, Frith MC. 2012. Adding unaligned sequences into an existing alignment using MAFFT and LAST. *Bioinformatics* **28:** 3144–3146. doi:10.1093/bioinformatics/bts578

Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C, et al. 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **28:** 1647–1649. doi:10.1093/bioinformatics/bts199

Kelly JA, Dinman JD. 2020. Structural and functional conservation of the programmed -1 ribosomal frameshift signal of SARS-CoV-2. bioRxiv doi:10.1101/2020.03.13.991083

Kim D, Lee JY, Yang JS, Kim JW, Kim VN, Chang H. 2020. The architecture of SARS-CoV-2 transcriptome. *Cell* **181:** 914–921.e10. doi:10.1016/j.cell.2020.04.011

Li W, Jaroszewski L, Godzik A. 2001. Clustering of highly homologous sequences to reduce the size of large protein databases. *Bioinformatics* **17:** 282–283. doi:10.1093/bioinformatics/17.3.282

Li L, Kang H, Liu P, Makkinje N, Williamson ST, Leibowitz JL, Giedroc DP. 2008. Structural lability in stem-loop 1 drives a 5′ UTR-3′ UTR interaction in coronavirus replication. *J Mol Biol* **377:** 790–803. doi:10.1016/j.jmb.2008.01.068

Libin PJK, Deforche K, Abecasis AB, Theys K. 2019. VIRULIGN: fast codon-correct alignment and annotation of viral genomes. *Bioinformatics* **35:** 1763–1765. doi:10.1093/bioinformatics/bty851

Liu P, Li L, Millership JJ, Kang H, Leibowitz JL, Giedroc DP. 2007. A U-turn motif-containing stem–loop in the coronavirus 5′ untranslated region plays a functional role in replication. *RNA* **13:** 763–780. doi:10.1261/rna.261807

Madhugiri R, Fricke M, Marz M, Ziebuhr J. 2014. RNA structure analysis of alphacoronavirus terminal genome regions. *Virus Res* **194:** 76–89. doi:10.1016/j.virusres.2014.10.001

Nawrocki EP, Eddy SR. 2013. Infernal 1.1: 100-fold faster RNA homology searches. *Bioinformatics* **29:** 2933–2935. doi:10.1093/bioinformatics/btt509

Plant EP, Dinman JD. 2008. The role of programmed-1 ribosomal frameshifting in coronavirus propagation. *Front Biosci* **13:** 4873–4881. doi:10.2741/3046

Rangan R, Watkins AM, Kladwang W, Das R. 2020a. De novo 3D models of SARS-CoV-2 RNA elements and small-molecule-binding RNAs to guide drug discovery. bioRxiv doi:10.1101/2020.04.14.041962

Rangan R, Zheludev IN, Das R. 2020b. RNA genome conservation and secondary structure in SARS-CoV-2 and SARS-related viruses. bioRxiv doi:10.1101/2020.03.27.012906

Rivas E, Clements J, Eddy SR. 2017. A statistical test for conserved RNA structure shows lack of evidence for structure in lncRNAs. *Nat Methods* **14:** 45–48. doi:10.1038/nmeth.4066

Rivas E, Clements J, Eddy SR. 2020. Estimating the power of sequence covariation for detecting conserved RNA structure. *Bioinformatics*. doi:10.1093/bioinformatics/btaa080

Robertson MP, Igel H, Baertsch R, Haussler D, Ares M Jr, Scott WG. 2005. The structure of a rigorously conserved RNA element within the SARS virus genome. *PLoS Biol* **3:** e5. doi:10.1371/journal.pbio.0030005

Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert M, Söding J, et al. 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol* **7:** 539. doi:10.1038/msb.2011.75

Spurgers KB, Sharkey CM, Warfield KL, Bavari S. 2008. Oligonucleotide antiviral therapeutics: antisense and RNA interference for highly pathogenic RNA viruses. *Antiviral Res* **78:** 26–36. doi:10.1016/j.antiviral.2007.12.008

Stammler SN, Cao S, Chen S-J, Giedroc DP. 2011. A conserved RNA pseudoknot in a putative molecular switch domain of the 3′-untranslated region of coronaviruses is only marginally stable. *RNA* **17:** 1747–1759. doi:10.1261/rna.2816711

Tzou PL, Huang X, Shafer RW. 2017. NucAmino: a nucleotide to amino acid alignment optimized for virus gene sequences. *BMC Bioinformatics* **18:** 138. doi:10.1186/s12859-017-1555-6

van den Born E, Posthuma CC, Gultyaev AP, Snijder EJ. 2005. Discontinuous subgenomic RNA synthesis in arteriviruses is guided by an RNA hairpin structure located in the genomic leader region. *J Virol* **79:** 6312–6324. doi:10.1128/JVI.79.10.6312-6324.2005

Viruses and Coronaviridae Study Group of the International Committee on Taxonomy. 2020. The species severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2. *Nat Microbiol* **5:** 536–544. doi:10.1038/s41564-020-0695-z

Washietl S, Hofacker IL. 2004. Consensus folding of aligned sequences as a new measure for the detection of functional RNAs by comparative genomics. *J Mol Biol* **342:** 19–30. doi:10.1016/j.jmb.2004.07.018

Wolfinger M. 2020. Evolutionarily conserved RNA structures in the upstream regions of Wuhan seafood market pneumonia virus (Wuhan-nCoV) and SARS virus. figshare. https://doi.org/10.6084/m9.figshare.11659575.v1

Wu F, Zhao S, Yu B, Chen Y-M, Wang W, Hu Y, Song Z-G, Tao Z-W, Tian J-H, Pei Y-Y, et al. 2020. A new coronavirus associated with human respiratory disease in China. *Nature* **579:** 265–269. doi:10.1038/s41586-020-2008-3

Yang D, Leibowitz JL. 2015. The structure and functions of coronavirus genomic 3′ and 5′ ends. *Virus Res* **206:** 120–133. doi:10.1016/j.virusres.2015.02.025

Yang D, Liu P, Wudeck EV, Giedroc DP, Leibowitz JL. 2015. SHAPE analysis of the RNA secondary structure of the mouse hepatitis virus 5′ untranslated region and N-terminal nsp1 coding sequences. *Virology* **475:** 15–27. doi:10.1016/j.virol.2014.11.001

Zhang K, Li S, Kappel K, Pintilie G, Su Z, Mou T-C, Schmid MF, Das R, Chiu W. 2019. Cryo-EM structure of a 40 kDa SAM-IV riboswitch RNA at 3.7 Å resolution. *Nat Commun* **10:** 5511. doi:10.1038/s41467-019-13494-7

RNA

A PUBLICATION OF THE RNA SOCIETY

RNA genome conservation and secondary structure in SARS-CoV-2 and SARS-related viruses: a first look

Ramya Rangan, Ivan N. Zheludev, Rachel J. Hagey, et al.

RNA 2020 26: 937-959 originally published online May 12, 2020
Access the most recent version at doi:10.1261/rna.076141.120