



# In-Depth Bioinformatic Analyses of *Nidovirales* Including Human SARS-CoV-2, SARS-CoV, MERS-CoV Viruses Suggest Important Roles of Non-canonical Nucleic Acid Structures in Their Lifecycles

## OPEN ACCESS

### Edited by:

Ilaria Frasson,  
University of Padova, Italy

### Reviewed by:

Naoki Sugimoto,  
Frontier Institute for Biomolecular  
Engineering Research, Japan  
Emanuela Ruggiero,  
University of Padova, Italy

### \*Correspondence:

Petr Pečinka  
petr.pecinka@osu.cz

†These authors have contributed  
equally to this work

### Specialty section:

This article was submitted to  
Antimicrobials, Resistance and  
Chemotherapy,  
a section of the journal  
Frontiers in Microbiology

Received: 13 April 2020

Accepted: 17 June 2020

Published: 03 July 2020

### Citation:

Bartas M, Brázda V, Bohálová N,  
Cantara A, Volná A, Stachurová T,  
Malachová K, Jagelská EB,  
Porubiaková O, Červeň J and  
Pečinka P (2020) In-Depth  
Bioinformatic Analyses of *Nidovirales*  
Including Human SARS-CoV-2,  
SARS-CoV, MERS-CoV Viruses  
Suggest Important Roles of  
Non-canonical Nucleic Acid  
Structures in Their Lifecycles.  
*Front. Microbiol.* 11:1583.  
doi: 10.3389/fmicb.2020.01583

Martin Bartas<sup>1†</sup>, Václav Brázda<sup>2,3†</sup>, Natálie Bohálová<sup>2,4</sup>, Alessio Cantara<sup>2,4</sup>, Adriana Volná<sup>5</sup>, Tereza Stachurová<sup>1</sup>, Kateřina Malachová<sup>1</sup>, Eva B. Jagelská<sup>2</sup>, Otilia Porubiaková<sup>2,3</sup>, Jiří Červeň<sup>1</sup> and Petr Pečinka<sup>1\*</sup>

<sup>1</sup> Department of Biology and Ecology, Faculty of Science, University of Ostrava, Ostrava, Czechia, <sup>2</sup> Department of Biophysical Chemistry and Molecular Oncology, Institute of Biophysics, Academy of Sciences of the Czech Republic, Brno, Czechia, <sup>3</sup> Faculty of Chemistry, Brno University of Technology, Brno, Czechia, <sup>4</sup> Department of Experimental Biology, Faculty of Science, Masaryk University, Brno, Czechia, <sup>5</sup> Department of Physics, Faculty of Science, University of Ostrava, Ostrava, Czechia

Non-canonical nucleic acid structures play important roles in the regulation of molecular processes. Considering the importance of the ongoing coronavirus crisis, we decided to evaluate genomes of all coronaviruses sequenced to date (stated more broadly, the order *Nidovirales*) to determine if they contain non-canonical nucleic acid structures. We discovered much evidence of putative G-quadruplex sites and even much more of inverted repeats (IRs) loci, which in fact are ubiquitous along the whole genomic sequence and indicate a possible mechanism for genomic RNA packaging. The most notable enrichment of IRs was found inside 5'UTR for IRs of size 12+ nucleotides, and the most notable enrichment of putative quadruplex sites (PQSS) was located before 3'UTR, inside 5'UTR, and before mRNA. This indicates crucial regulatory roles for both IRs and PQSS. Moreover, we found multiple G-quadruplex binding motifs in human proteins having potential for binding of SARS-CoV-2 RNA. Non-canonical nucleic acid structures in *Nidovirales* and in novel SARS-CoV-2 are therefore promising druggable structures that can be targeted and utilized in the future.

**Keywords:** coronavirus, genome, RNA, G-quadruplex, inverted repeats

## INTRODUCTION

The order *Nidovirales* is a monophyletic group of animal RNA viruses. This group can be divided into the six distinct families of *Arteriviridae*, *Coronaviridae*, *Mesnidovirineae*, *Mononiviridae*, *Ronidovirineae*, and *Tobaniviridae*. All known *Nidovirales* have single-stranded, polycistronic RNA genomes of positive polarity (Modrow et al., 2013). Due to the Severe acute respiratory syndrome-related coronavirus (SARS-CoV) epidemic (November 2002–July 2003, Southern China), Middle East respiratory syndrome-related coronavirus (MERS-CoV) outbreaks (January 2014–May 2014,

Saudi Arabia and May 2015–June 2015, South Korea), and the most recent Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) worldwide pandemic (starting in November 2019 in Wuhan, China), this viral group is now extensively studied (Hung, 2003; Cowling et al., 2015; Oboho et al., 2015; Li et al., 2020).

The small size of RNA virus genomes is in principle linked to their limited ability for RNA synthesis, which is directly connected to a replication complex containing RNA-dependent RNA polymerase (RdRp) without repair mechanisms (Gorbalenya et al., 2006). These RNA viruses can generate one mutation per genome per replication round (Drake and Holland, 1999). This combination of features means that RNA viruses are able to adapt to new environmental conditions, but they are limited in expanding their genomes because they must keep their mutation load low so that their survival is possible (Eigen and Schuster, 1977; Lauring et al., 2013; Carrasco-Hernandez et al., 2017). *Nidovirales* bind to their host cell receptors on the cell surface, after which fusion of the viral and cellular membranes is mediated by one of the surface glycoproteins. This mechanism, which progresses in the cytoplasm or endosomal membrane, releases the nucleocapsid into the host cell's cytoplasm. After genome "transportation," translation of two replicase open reading frames (ORFs) is initiated by host ribosomes. This results in large polyprotein precursors that undergo autoproteolysis to eventually produce a membrane-bound replicase/transcriptase complex. This complex initiates synthesis of the genome RNA and controls the synthesis of structural and some other proteins. New virus particles are assembled by association of the new genomes with the cytoplasmic nucleocapsid protein and subsequent envelopment of the nucleocapsid structure. Subsequently, the viral envelope proteins are inserted into intracellular membranes and targeted to the site of virus assembly (most often membranes between the endoplasmic reticulum and Golgi complex) and then they meet up with the nucleocapsid and trigger the budding of virus particles into the lumen of the membrane compartment. The newly formed virions then leave the cell by following the exocytic pathway toward the plasma membrane (Lai and Cavanagh, 1997; Gorbalenya, 2001; Snijder et al., 2003; Ziebuhr, 2004; Gorbalenya et al., 2006).

Today, SARS-CoV-2 is being studied intensively by scientists all over the world due to the ongoing 2020 coronavirus pandemic (Cohen, 2020). The origin of this virus is unknown, but recently a two-hit scenario was proposed wherein SARS-CoV-2 ancestors in bats first acquired genetic characteristics of SARS-CoV by incorporating a SARS-like receptor-binding domain through recombination prior to 2009; subsequently, the lineage that led to SARS-CoV-2 accumulated further, unique changes specifically within this domain (Patino-Galindo et al., 2020). As true of SARS-CoV, cell entry by SARS-CoV-2 depends upon angiotensin-converting enzyme 2 (ACE2) and transmembrane protease, serine 2 (TMPRSS2) for viral spike protein priming (Hoffmann et al., 2020). The genome of SARS-CoV-2 was sequenced and annotated in early January 2020 (Wu et al., 2020). A recent study revealed that the transcriptome of SARS-CoV-2 is highly complex due to numerous canonical and non-canonical

recombination events. Moreover, it was found that SARS-CoV-2 produces transcripts encoding unknown ORFs and at least 41 potential RNA modification sites with an AAGAA motif were discovered in its RNA (Kim et al., 2020).

G-quadruplex binding proteins (QBPs) play crucial roles in many signaling pathways, including such biologically highly relevant activities as cell division, dysregulations of which lead to cancer development (Wu et al., 2008; Brázda et al., 2014). QBPs have been found to be involved in various viral infection pathways. An interesting example is HIV-1 nucleocapsid protein NCp7. It has been described how this protein helps to resolve an otherwise very stable G-quadruplex structure in viral RNA that stalls reverse transcription (Butovskaya et al., 2019). Various G-quadruplex-forming aptamers are used as drugs against many different viral proteins, suggesting a prominent role for QBP-mediated regulation. Among many other viruses, in particular Hepatitis C virus, HIV-1, and SARS-CoV are targets of these G-quadruplex-forming aptamers (Platella et al., 2017). Quadruplex binding domain has been found in non-structural protein 3 (Nsp3) (Lei et al., 2018). This so-called SARS-UNIQUE-DOMAIN (SUD), and especially its M subdomain, was observed to be essential for SARS replication in host cells. Deletion or even substitution mutations in key RNA-interacting amino acids were shown to result in viral inability to replicate within host cells (Kusov et al., 2015). Moreover, this subdomain was found also in MERS and several other coronaviruses (Kusov et al., 2015).

G-quadruplexes are secondary nucleic acid structures formed in guanine-rich strands (Burge et al., 2006; Vorlíčková et al., 2012; Kolesnikova and Curtis, 2019). These have been detected in various genomes, but most extensively they have been described in human genomes (Chambers et al., 2015; Bedrat et al., 2016; Hänsel-Hertsch et al., 2016). They are present also in viruses (Lavezzo et al., 2018; Frasson et al., 2019). G-quadruplexes probably play an important role in regulating replication in most viral nucleic acids (Lavezzo et al., 2018), and these structures have been suggested as targets for antiviral therapy (Métifiot et al., 2014; Ruggiero and Richter, 2018, 2020). Along with cruciforms and hairpins, which can be formed in nucleic acids by inverted repeats (IRs), G-quadruplexes are significant genome elements playing specific biological roles. They are involved, for example, in the regulation of DNA replication and transcription (Bagga et al., 1990; Yu, 2009; Brázda et al., 2011). It has been demonstrated that IRs are important for various processes in viruses, including their genome organization (Li and Li, 2010; Ishimaru et al., 2013; Xie et al., 2017; Bridges et al., 2019). Another interesting RNA motif that has been used as a drug target and was found in SARS-CoV targeted by 1,4-diazepam is the "slippery sequence" followed by a pseudoknot (Plant et al., 2005). This structure, common among all coronaviruses, works based on ribosomal  $-1$  frameshifting that switches on viral fusion proteins (Plant et al., 2005).

In all prokaryotic and eukaryotic cells, as well as viruses, there have been found sequence motifs such as IR sequences forming cruciforms and hairpins or G-rich sequences that form G-quadruplexes (Brázda et al., 2011; Lavezzo et al., 2018; Bartas et al., 2019). In the present research, we conducted a systematic and comprehensive bioinformatic study searching

for the occurrence of IRs and putative quadruplex sites (PQSs) within the genomes of all known *Nidovirales*. The aim was to find one or more potential druggable RNA targets to address the present COVID-19 threat (Figure 1).

## MATERIALS AND METHODS

### Genome Source

Linear genomes of 109 viruses from the order *Nidovirales* were downloaded from the genome database of the National Center for Biotechnology Information (NCBI) (Federhen, 2011). Full names, phylogenetic groups, exact NCBI accession numbers, and further information are summarized in **Supplementary Material 1**.

### Nidovirales Phylogenetic Tree Construction

Exact taxonomic identifiers of all analyzed *Nidovirales* species (obtained from Taxonomy Browser via NCBI Taxonomy Database Drake and Holland, 1999) were downloaded to phyloT: a tree generator (<http://phylo.t.biobyte.de>) and a phylogenetic tree was constructed using the function “Visualize in iTOL” in the Interactive Tree of Life environment (Letunic and Bork, 2019).

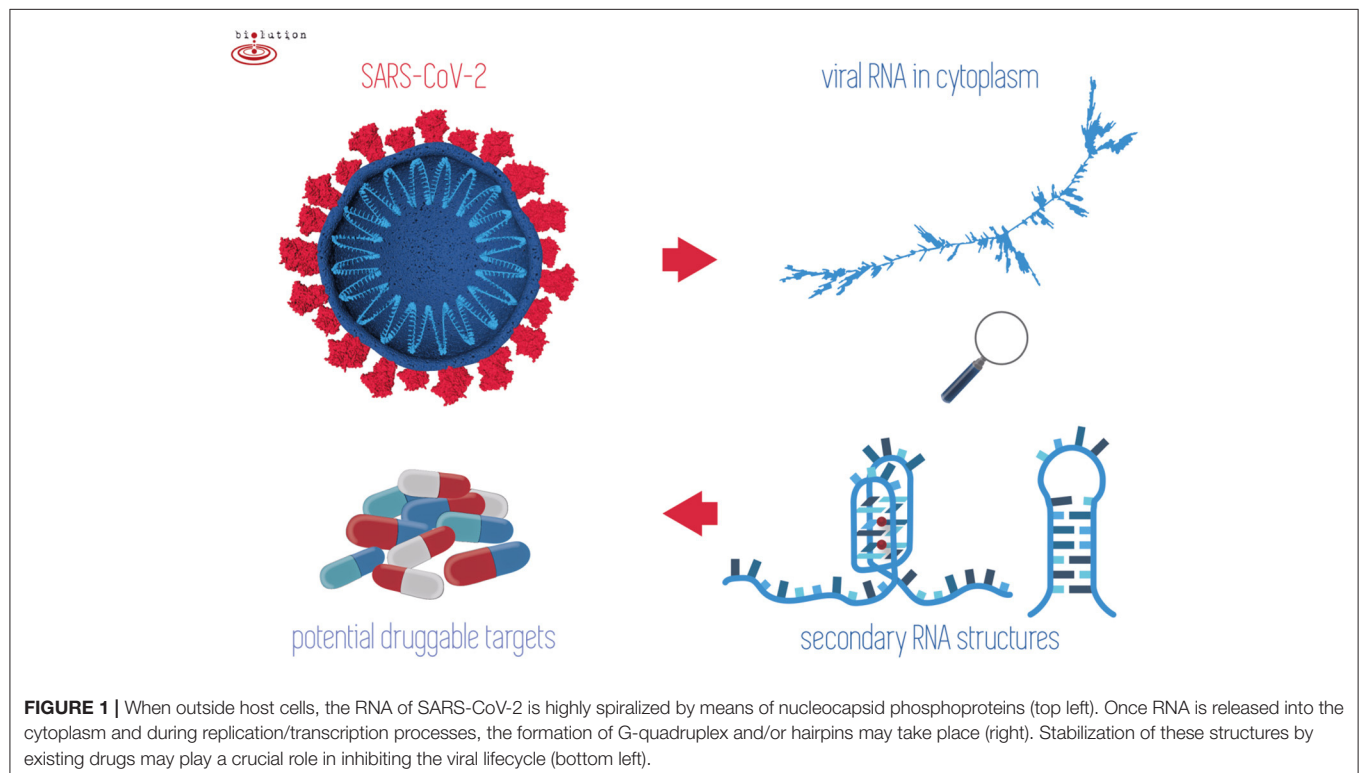
### Analyses of PQSs Occurrence in *Nidovirales* Sequences

*Nidovirales* sequences were analyzed using our G4Hunter Web Tool (Brázda et al., 2019), and selected sequences were verified also by QGRS mapper (Kikin et al., 2006) and G4screener (Garant et al., 2018). G4Hunter web is a more

recent tool compared to Quadparser and QGRS mapper, and this algorithm allows quantitative analyses whereby G4 propensity is calculated depending on G richness and G skewness. Moreover, a new implementation of G4Hunter allows for performing batch analyses. In general, the earlier algorithms did not consider atypical quadruplex-forming structures, as have been described in various outstanding articles (Guedin et al., 2010; Kocman and Plavec, 2017; Lightfoot et al., 2019). The G4Hunter Web software is capable to read the NCBI identifier of the sequences uploaded in a.csv file. The parameters for G4Hunter were set to 25 as window size and G4Hunter score above 1.2. The results for each analyzed sequence contained information about the size of the genome and number of putative PQSs. All the results were merged into a single Microsoft Excel file where statistical analysis was then made. We also downloaded features tables of each virus from the NCBI database. These tables contain information about known features in the genome of each species. We searched the occurrence of G-quadruplex-forming sequences before, inside, and after the specific features of each genome using a predefined feature neighborhood  $\pm 100$  nucleotides (nt). Data were then plotted in Excel, where the statistical analysis was made. Complete analyses of PQS occurrence in *Nidovirales*, including a list of those PQSs found, are provided in **Supplementary Material 2**.

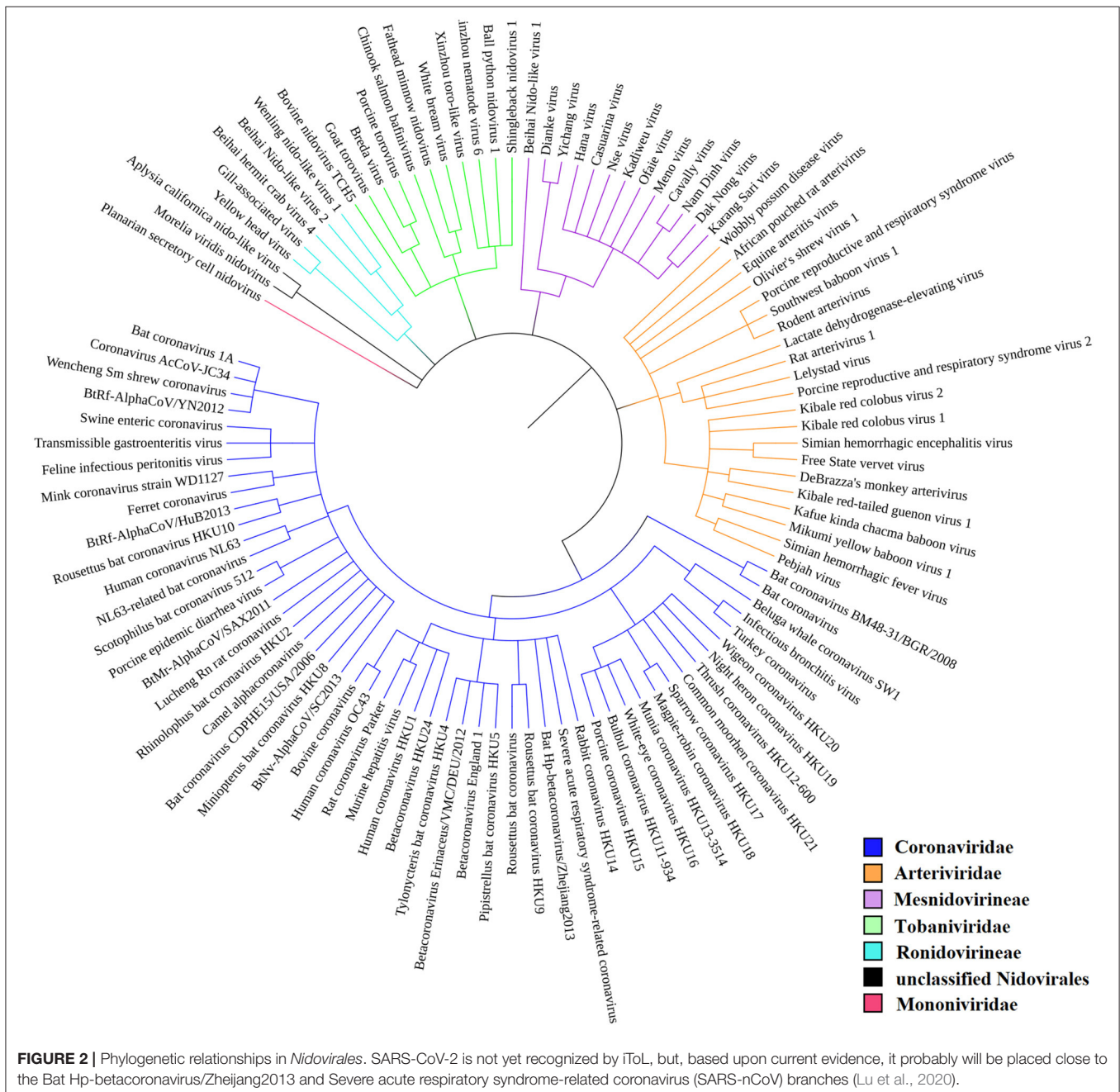
### Analyses of IRs Occurrence in *Nidovirales* Sequences

All *Nidovirales* genomes were analyzed by the core of our Palindrome analyzer webserver (Brázda et al., 2016). The



software was modified to read NCBI identifiers of sequences from text files. The size of IRs was set to 6–30 nt, size of spacers to 0–10 nt, and maximally one mismatch was allowed. A separate list of IRs in each of the 109 sequences and an overall report were exported. The overall report contained the lengths of the analyzed sequences, total number of IRs found, numbers of IRs grouped by size of IR (6–30 nt), and sum of IRs longer than 8, 10, and 12 nt. The software also counted frequency of IRs in each sequence. Frequencies of IRs were normalized per 1,000 nt. Features tables of 109 *Nidovirales* genomes were downloaded from the NCBI database and grouped

by their names as stated in the feature table file. Analyses of IRs occurrence inside and around (before and after) these features was performed. The search for IRs took place in predefined feature neighborhoods  $\pm 100$  nt around and inside feature boundaries. We calculated the numbers of all IRs and of those longer than 8, 10, and 12 nt in regions before, inside, and after the features. The categorization of an IR according to its overlap with a feature or feature neighborhood is demonstrated by the example shown in **Supplementary Material 3**. Complete analyses of IRs occurrence in *Nidovirales* are provided in **Supplementary Material 4**.



## RNA Fold Predictions

In order to be able to display higher structures of the coronavirus genome, we used Galaxy's free-online webserver (Afgan et al., 2018) and its RNA fold tool (Lorenz et al., 2011). This tool allows quick calculation of minimum free energy of secondary structures. We left the default parameters (Temperature 37°C, Unpaired bases to participate in all dangling ends, Naview layout). We used SARS-CoV-2 genomic RNA sequence (NC\_045512.2) in FASTA format as the input format. The output data were then displayed using the RNA plot tool (Lorenz et al., 2011). We again left the default parameters (Naview layout, Output format Postscript.ps). We then downloaded the displayed secondary structures in high-resolution format. The raw data are provided in **Supplementary Material 5**.

## Multiple Alignment of SUD Domains (M Regions) in Nsp3 of Pathogenic Species

Multiple protein alignment was done using MUSCLE (Edgar, 2004) under default parameters [UGENE (Okonechnikov et al., 2012) workflow was used]. The following accessions were used: NP\_828862.2 (Nsp3 SARS-CoV), YP\_009047231.1 (Nsp3 MERS-CoV), and YP\_009725299.1 (Nsp3 SARS-CoV-2). Conserved regions were added according to a graph published previously (Kusov et al., 2015).

## Prediction of Human RNA-Binding Protein Sites in SARS-CoV-2 RNA

The human SARS-CoV-2 RNA sequence was downloaded from NCBI (accession NC\_045512.2) in FASTA format and inserted into the RBPmap (Mapping Binding Sites of RNA-binding proteins) web-based tool (Paz et al., 2014). The database of 114 human experimentally validated motifs was used. Both the "High stringency" and "Apply conservation filter" options were used. The output was further filtered in Excel to keep only those hits below  $p = 1.10^{-6}$ . The complete results are provided in **Supplementary Material 6**.

## Statistical Analyses

A cluster dendrogram of PQS characteristics was constructed in the program R, version 3.6.3, using the *pvclust* package to further reveal and graphically depict similarities between particular *Nidovirales* species (**Figure 4**). The following values were used as input data: frequency of PQS per 1,000 nt with threshold G4Hunter score of 1.2; frequency of PQS per 1,000 nt with threshold G4Hunter score of 1.4; and % PQS in genome (coverage) (**Supplementary Material 7**). A cluster dendrogram of IRs (**Figure 6**) was constructed from these values: frequency of IRs per 1,000 nt; frequency of IRs per 1,000 nt of length 8+; frequency of IRs per 1,000 nt of length 10+; and frequency of IRs per 1,000 nt of length 12+. The following parameters were used for both analyses: cluster method "ward.D2," distance "euclidean," number of bootstrap resampling was 10,000. Statistically significant clusters (based on AU values above 95, equivalent to  $p < 0.05$ ) are highlighted by rectangles marked with broken red lines. R code is provided in **Supplementary Material 7**.

To determine whether SARS-CoV-2 significantly differs in frequencies of PQSs and IRs compared to randomly shuffled sequences ( $N = 10$ ) of the same length and nucleotide composition, we performed a two-sided Wilcoxon signed-rank test (with  $p$ -value cutoff at 0.05). This test was run for plus and minus strands separately.

## RESULTS

### Phylogenetic Relationships in *Nidovirales*

According to the current state of knowledge, the order *Nidovirales* can be divided into six distinct families (Liò and Goldman, 2004; Hanada et al., 2005; Chen et al., 2020). There are just two unclassified species among all those species recorded within NCBI Genome (**Figure 2**). *Coronaviridae* is the largest family and consists of 56 species. Among these are 12 species able

**TABLE 1** | Genomic sequence sizes, PQS frequencies, and total counts.

| All                            | Seq | Median | Short  | Long   | PQS   | Mean f | Min f | Max f | GC%   |
|--------------------------------|-----|--------|--------|--------|-------|--------|-------|-------|-------|
| Nidovirales                    | 109 | 26,975 | 12,704 | 41,178 | 1,021 | 0.48   | 0     | 9.07  | 42.15 |
| <b>Family</b>                  |     |        |        |        |       |        |       |       |       |
| Arteriviridae                  | 22  | 15,236 | 12,704 | 15,728 | 392   | 1.18   | 0.45  | 2.72  | 51.60 |
| Coronaviridae                  | 56  | 28,345 | 20,398 | 31,686 | 162   | 0.10   | 0     | 0.52  | 39.39 |
| Mesnidovirineae                | 12  | 20,117 | 19,867 | 20,949 | 250   | 1.03   | 0     | 9.07  | 38.75 |
| Ronidovirineae                 | 5   | 26,253 | 24,648 | 29,385 | 51    | 0.39   | 0.04  | 1.07  | 44.79 |
| Tobaniviridae                  | 11  | 27,318 | 20,261 | 33,452 | 129   | 0.43   | 0     | 1.48  | 40.94 |
| <b>By hosts</b>                |     |        |        |        |       |        |       |       |       |
| Invertebrates                  | 17  | 20,307 | 19,917 | 41,178 | 298   | 0.82   | 0     | 9.07  | 39.76 |
| Vertebrates (including Humans) | 83  | 27,608 | 12,704 | 33,452 | 591   | 0.38   | 0     | 1.75  | 42.74 |
| Humans                         | 13  | 29,751 | 27,317 | 31,028 | 21    | 0.06   | 0     | 0.32  | 37.88 |
| SARS-CoV-2                     | 1   | 29,903 | –      | –      | 1     | 0.03   | –     | –     | 37.97 |

Seq (total number of sequences), Median (median length of sequences), Short (shortest sequence), Long (longest sequence), PQS (total number of predicted PQSs), Mean f (mean frequency of predicted PQS per 1,000 nt), Min f (lowest frequency of predicted PQS per 1,000 nt), Max f (highest frequency of predicted PQS per 1,000 nt), GC% (mean GC content). Note that 132 PQSs were in *Nidovirales* with Undefined host.

to infect humans, including SARS-CoV, MERS-CoV, and SARS-CoV-2. Second largest is the *Arteriviridae* family with 22 species, third largest is the *Mesnidovirineae* family with 13 species, fourth is the *Tobaniviridae* family with 11 species, including 1 species able to infect humans. Fifth largest is the *Ronidovirineae* family with 5 species, and sixth is the *Mononiviridae* family with only 1 species.

## Variation in PQS Frequency in Nidovirales

We analyzed the occurrence of PQSs using G4Hunter in all 109 known genomes of *Nidovirales*. The length of genomes in the dataset varies from 12,704 nt (Equine arteritis virus) to 41,178 nt (Planarian secretory cell nidovirus). The mean GC content is 42.15%, the minimum is 27.50% for Planarian secretory cell nidovirus (*Mononiviridae* family), and the maximum is 57.35% for Beihai Nido-like virus 1 (*Mesnidovirineae* family). Using standard values for the G4Hunter algorithm (i.e., window size 25 and G4Hunter score above 1.2), we found 1,021 PQSs among all 109 *Nidovirales* genomes. The most abundant PQSs are those with G4Hunter scores of 1.2–1.4 (98.24% of all PQS), much less abundant are PQSs with G4Hunter scores 1.4–1.6 (1.76% of all PQS), and there were no PQSs above the 1.6 G4Hunter score threshold. In general, a higher G4Hunter score means a higher probability of G-quadruplexes forming inside the PQS (Bedrat et al., 2016). Genomic sequence sizes, total PQS counts, and PQS frequencies characteristics are summarized in **Table 1**.

The highest PQS frequencies were found in Beihai Nido-like virus 1, which is an intracellular parasite of sea snails in genus *Turritella*. Beihai Nido-like virus 1 has in total 184 PQSs in its genomic sequence 20,278 nt long and PQS frequency of 9.07 PQS per 1,000 nt. On the other hand, no PQSs were found in the genomic sequences of 16 other *Nidovirales* species. The mean PQS value for all *Nidovirales* was 0.48 PQS per 1,000 nt. By viral families, the highest mean PQS frequency per 1,000 nt was in *Arteriviridae* (1.18), followed by in *Mesnidovirineae* (1.03), much lower in *Tobaniviridae* (0.43) and in *Ronidovirineae* (0.39), and the lowest was in *Coronaviridae* (0.10). When grouped and analyzed by host organisms, new information became apparent. The highest mean PQS frequency was in *Nidovirales* infecting Invertebrates (0.80), then in *Nidovirales* infecting Vertebrates including Humans (0.38), and the last in *Nidovirales* infecting Humans (0.06). In pathogenic human coronaviruses, the total PQSs counts were as follow: 4 PQS in SARS-CoV, 0 PQS in MERS-CoV, and 1 PQS in SARS-CoV-2. To test whether the PQS frequency per 1,000 nt in SARS-CoV-2 is significantly different than in a randomly shuffled sequence, we generated 10 scrambled sequences with length and nucleotide content the same as in SARS-CoV-2. In that test, the mean PQS frequency per 1,000 nt was 0.191 and standard deviation was 0.101. We performed Wilcoxon signed-rank test with continuity correction and the resulting *p*-value was 0.008. Thus, the PQS frequency per 1,000 nt in SARS-CoV-2 (0.033) is significantly lower than expected (only about one-sixth).

All PQSs found in ranges of G4Hunter score intervals and precomputed PQS frequencies per 1,000 nt are summarized in **Table 2**. The relationships between observed PQS frequency per 1,000 nucleotides and GC content in all analyzed *Nidovirales*

**TABLE 2 |** Total number of PQSs and their resulting frequencies per 1,000 nt in all 109 genomes of *Nidovirales* and in particular categories according to their hosts, grouped by G4Hunter score.

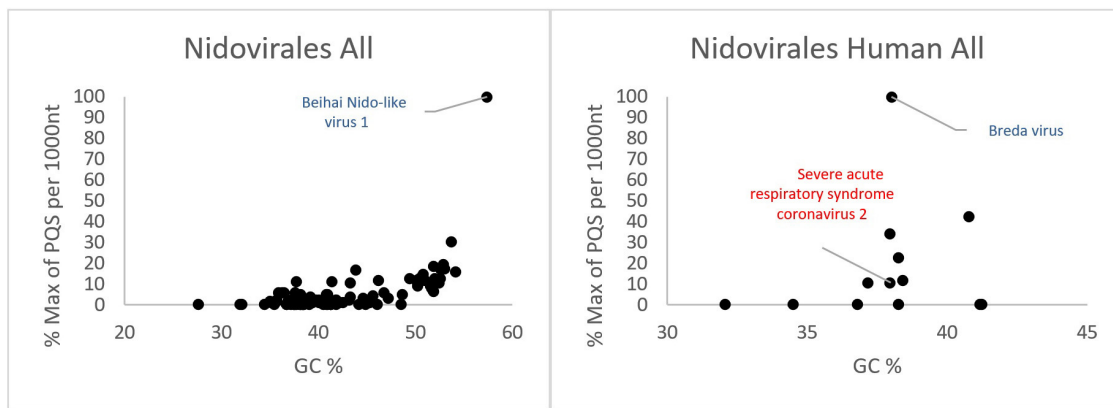
| Interval of G4Hunter Score           | Number of PQSs in dataset | PQS frequency per 1,000 nt |
|--------------------------------------|---------------------------|----------------------------|
| <b>All</b>                           |                           |                            |
| 1.2                                  | 1,003                     | 0.47                       |
| 1.4                                  | 18                        | 0.01                       |
| <b>Arteriviridae</b>                 |                           |                            |
| 1.2                                  | 386                       | 1.17                       |
| 1.4                                  | 6                         | 0.02                       |
| <b>Coronaviridae</b>                 |                           |                            |
| 1.2                                  | 157                       | 0.10                       |
| 1.4                                  | 5                         | 0.003                      |
| <b>Mesnidovirineae</b>               |                           |                            |
| 1.2                                  | 245                       | 1.01                       |
| 1.4                                  | 5                         | 0.02                       |
| <b>Ronidovirineae</b>                |                           |                            |
| 1.2                                  | 50                        | 0.38                       |
| 1.4                                  | 1                         | 0.01                       |
| <b>Tobaniviridae</b>                 |                           |                            |
| 1.2                                  | 129                       | 0.43                       |
| 1.4                                  | 0                         | 0                          |
| <b>Humans only</b>                   |                           |                            |
| 1.2                                  | 20                        | 0.05                       |
| 1.4                                  | 1                         | 0.003                      |
| <b>Vertebrate (Including humans)</b> |                           |                            |
| 1.2                                  | 580                       | 0.37                       |
| 1.4                                  | 11                        | 0.01                       |
| <b>Invertebrates</b>                 |                           |                            |
| 1.2                                  | 292                       | 0.80                       |
| 1.4                                  | 6                         | 0.02                       |

Frequency was computed using total number of PQSs in each category divided by total length of all analyzed sequences and multiplied by 1,000. Note that 132 PQSs were in *Nidovirales* with Undefined host.

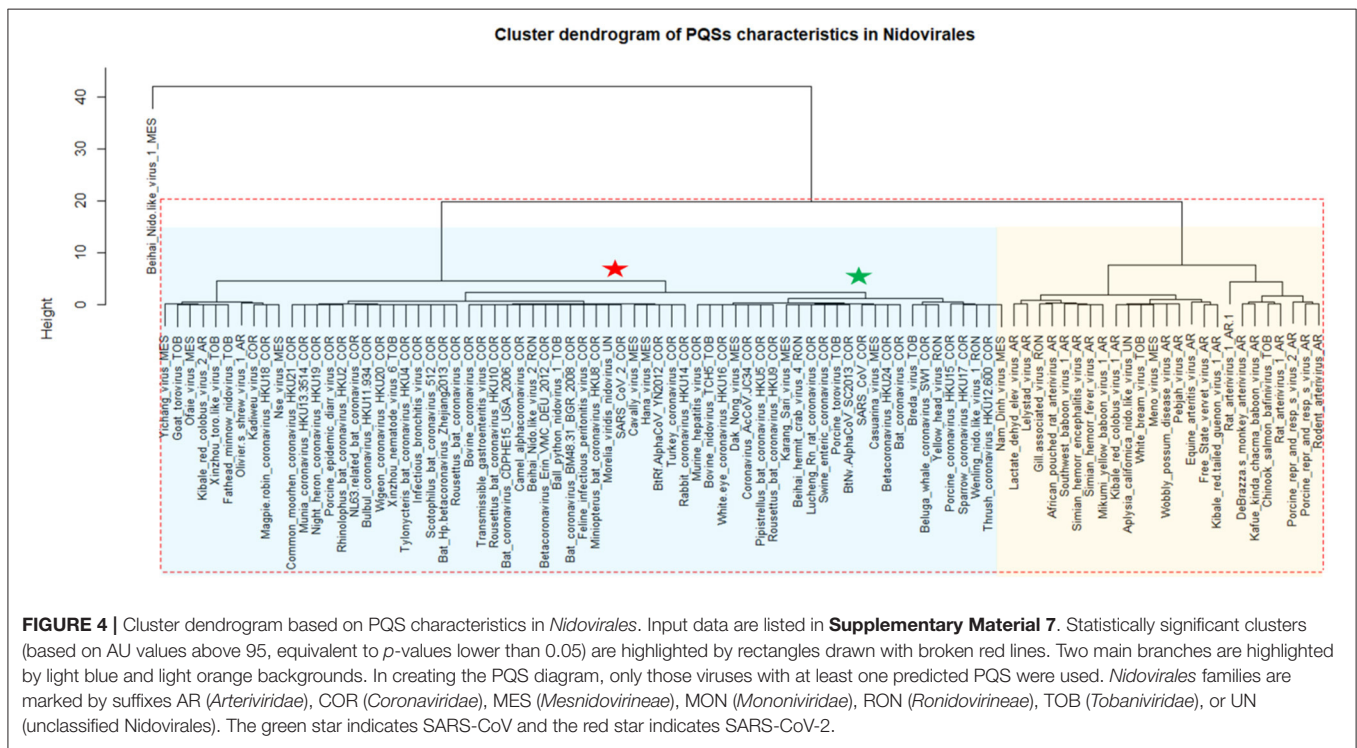
sequences are depicted in **Figure 3**. The Beihai Nido-like virus has both the highest GC content and highest PQS frequency. Within *Nidovirales* with Humans as host, however, the highest PQS frequency was found in Breda virus, which does not have the highest GC content in the dataset.

Most of the PQSs have G4Hunter scores between 1.2 and 1.4. Only a few sequences have G4Hunter scores above 1.4. In comparison with other *Nidovirales*, the members of the *Coronaviridae* and especially pathogenic human *Coronaviridae* have the lowest PQS frequency in the dataset.

A cluster dendrogram based on the PQS characteristics (**Figure 4**) further revealed interesting information. Beihai Nido-like virus 1 was confirmed as an outlier from the rest of the *Nidovirales*. All viral species belonging to the largest *Coronaviridae* family are located together in the middle of the cluster dendrogram. The relatively abundant family *Arteriviridae* clustered together mainly on the right side of the cluster dendrogram. Based



**FIGURE 3 |** Relationship between observed PQS frequency per 1,000 nucleotides and GC content in all analyzed *Nidovirales* sequences. In each G4Hunter score interval miniplot, frequencies were normalized according to the highest observed PQS frequency. Organisms with maximum frequency per 1,000 nt >50% are described and highlighted in color.



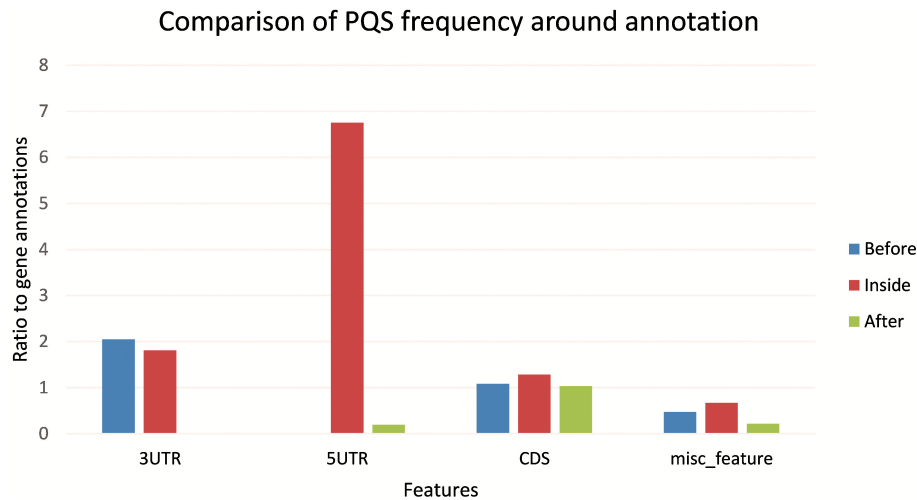
**FIGURE 4 |** Cluster dendrogram based on PQS characteristics in *Nidovirales*. Input data are listed in **Supplementary Material 7**. Statistically significant clusters (based on AU values above 95, equivalent to *p*-values lower than 0.05) are highlighted by rectangles drawn with broken red lines. Two main branches are highlighted by light blue and light orange backgrounds. In creating the PQS diagram, only those viruses with at least one predicted PQS were used. *Nidovirales* families are marked by suffixes AR (*Arteriviridae*), COR (*Coronaviridae*), MES (*Mesnidovirineae*), MON (*Mononiviridae*), RON (*Ronidovirineae*), TOB (*Tobaniviridae*), or UN (unclassified *Nidovirales*). The green star indicates SARS-CoV and the red star indicates SARS-CoV-2.

on dendrogram distances, we highlighted two main branches—one consisting mainly of *Coronaviridae* (light blue background) and the second of *Arteriviridae* (light orange background).

**Figure 5** shows the differences in PQS frequency by annotated loci. We downloaded features for every virus genome and analyzed the presence of PQS in each annotated sequence and in its proximity (before and after). The most notable enrichment of PQSs was located inside 5'UTR and before and inside 3'UTR. The lowest PQS frequencies were found after 3'UTR, before 5'UTR, and after miscellaneous features.

### Variation in Frequency of Inverted Repeats in Nidovirales

To find short IRs in *Nidovirales* genomic sequences, we utilized the core of the *Palindrome analyzer* (Brázda et al., 2016). We used values for *Palindrome analyzer* returning inverted repeats capable to form stable hairpins (size of IRs was set to 6–30 nt, size of spacer to 0–10 nt, maximally one mismatch). Genomic sequence sizes, total IR counts, and IR frequencies characteristics are summarized in **Table 3**. In total, 93,369 IRs were found and the mean IRs frequency was 33.90 per 1,000 nt. The maximal IR frequency was found in Planarian secretory cell nidovirus (56.17),



**FIGURE 5** | Differences in PQS frequency by annotated locus. The figure shows the PQS frequencies between annotations from the NCBI database. We analyzed frequencies of all PQSs inside, before, and after the annotations.

**TABLE 3** | Genomic sequence sizes, IRs frequencies and total counts.

| Order                          | Seq | Median | Short  | Long   | IRs    | Mean f | Min f | Max f | GC%   |
|--------------------------------|-----|--------|--------|--------|--------|--------|-------|-------|-------|
| Nidovirales                    | 109 | 26,975 | 12,704 | 41,178 | 93,369 | 33.90  | 19.23 | 56.17 | 42.15 |
| <b>Family</b>                  |     |        |        |        |        |        |       |       |       |
| Arteriviridae                  | 22  | 15,236 | 12,704 | 15,728 | 9,695  | 29.44  | 26.47 | 33.57 | 51.60 |
| Coronaviridae                  | 56  | 28,345 | 20,398 | 31,686 | 58,943 | 37.08  | 31.13 | 48.16 | 39.39 |
| Mesnidovirineae                | 12  | 20,117 | 19,867 | 20,949 | 8,012  | 33.04  | 19.23 | 39.06 | 38.75 |
| Ronidovirineae                 | 5   | 26,253 | 24,648 | 29,385 | 3,474  | 26.37  | 22.20 | 29.62 | 44.79 |
| Tobaniviridae                  | 11  | 27,318 | 20,261 | 33,452 | 9,082  | 30.24  | 23.96 | 37.95 | 40.94 |
| <b>Host</b>                    |     |        |        |        |        |        |       |       |       |
| Invertebrates                  | 17  | 20,307 | 19,917 | 4,178  | 13,233 | 31.77  | 19.23 | 56.17 | 39.76 |
| Vertebrates (including Humans) | 83  | 27,608 | 12,704 | 33,452 | 45,844 | 34.25  | 24.18 | 48.16 | 42.74 |
| Humans                         | 13  | 29,751 | 27,317 | 31,028 | 14,408 | 38.13  | 33.89 | 43.44 | 37.88 |
| SARS-CoV-2                     | 1   | 29,903 | –      | –      | 1,203  | 40.23  | –     | –     | 37.97 |

Seq (total number of sequences), Median (median length of sequences), Short (shortest sequence), Long (longest sequence), GC% (mean GC content), IRs (total number of predicted IRs), Mean f (mean frequency of predicted IRs per 1,000 nt), Min f (lowest frequency of predicted IRs per 1,000 nt), and Max f (highest frequency of predicted PQS per 1,000 nt) are shown for all 109 Nidovirales sequences, families, host groups, and for SARS-CoV-2 alone.

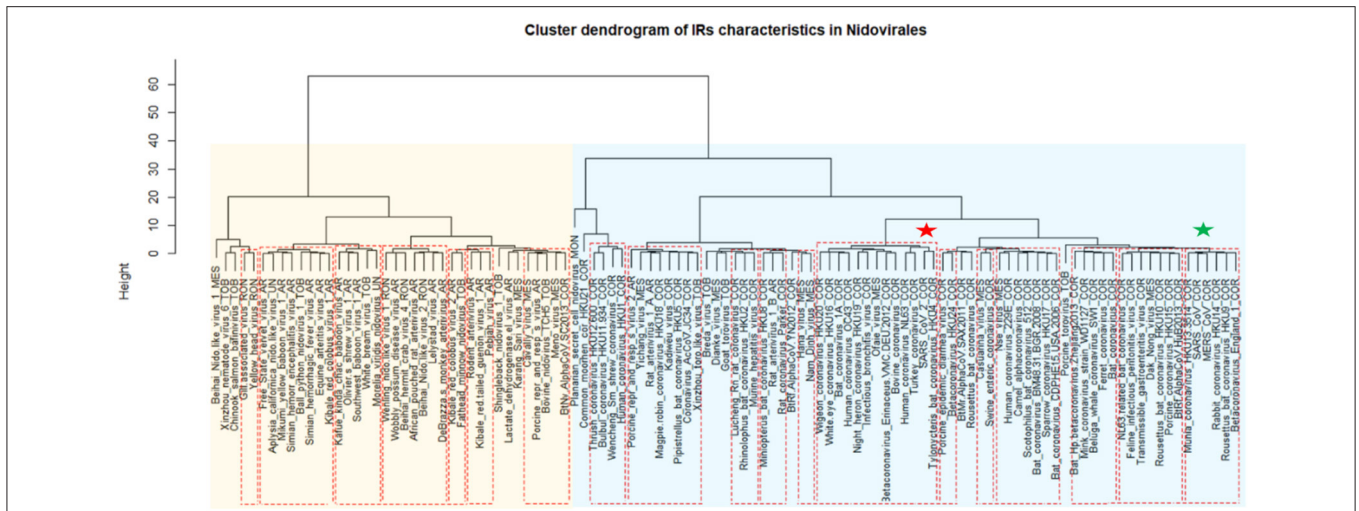
the species with the largest genome among *Nidovirales*. It has been suggested that Planarian secretory cell nidovirus diverged early from multi-ORF *Nidovirales* and acquired additional genes, including those typical of large DNA viruses or hosts (RNase T2, Ankyrin, and Fibronectin type II), which might modulate virus–host interactions (Saber et al., 2018). The lowest IR frequency was noticed in Beihai Nido-like virus 1 (19.23). Noteworthy is that this is the species of *Nidovirales* with the highest GC content and PQS frequency.

A cluster dendrogram based on the IR characteristics (Figure 6) further revealed significant differences in IR presence within *Nidovirales* genomes. Based on dendrogram distances, we highlighted two main branches—one consisting mainly of *Coronaviridae* (light blue background) and the second of *Arteriviridae* (light orange background). It is noteworthy that

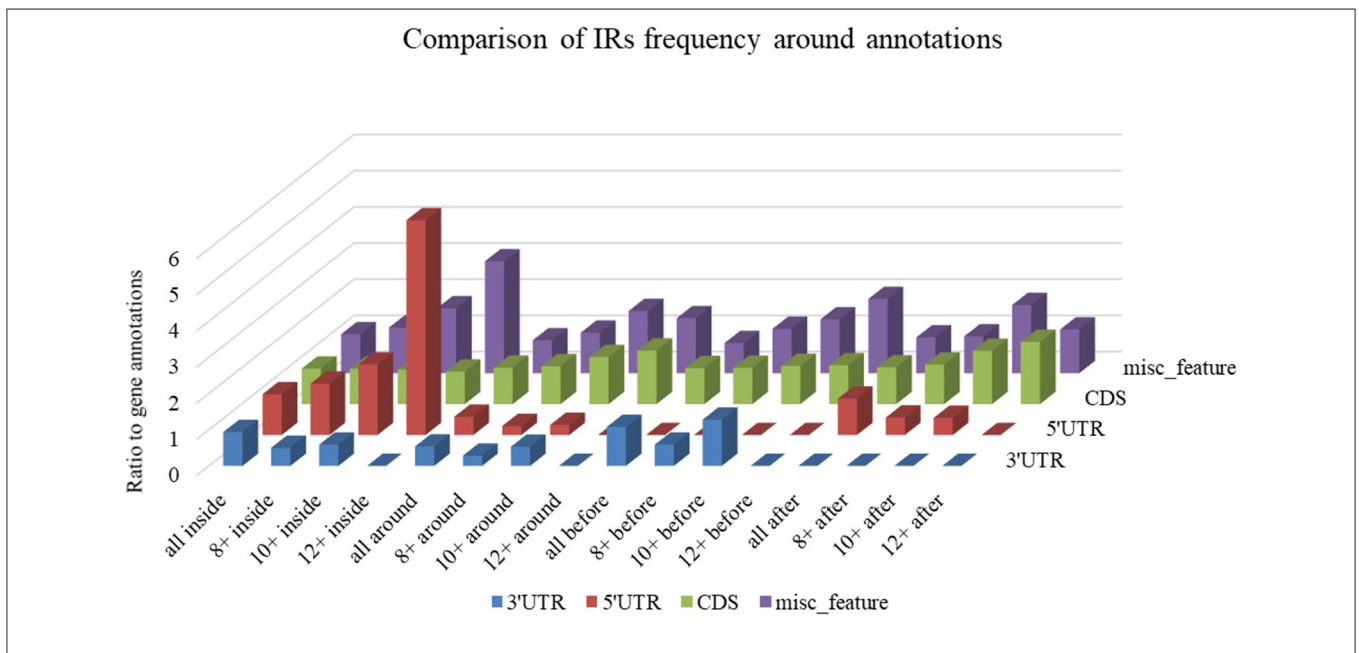
SARS-CoV and MERS-CoV are clustered adjacent to one another (green asterisk), but the novel coronavirus SARS-CoV-2 is located relatively far away from them (red asterisk).

Differences in IR frequency by annotated locus are depicted in Figure 7. By families, the highest mean IRs frequency was found in *Coronaviridae* (37.08) and lowest in *Ronidovirineae* (26.37). Novel human coronavirus SARS-CoV-2 has relatively high IR frequency per 1,000 nt in comparison with other *Nidovirales*. To test if the frequency of IRs per 1,000 nt in SARS-CoV-2 is significantly greater than in randomly shuffled sequences, we generated 10 scrambled sequences with length and nucleotide content the same as in SARS-CoV-2 (for both positive and negative strands). For both positive and negative strands, the results were very similar (mean frequencies of IRs per 1,000 nt were  $34.43 \pm 0.54$  and  $34.47$





**FIGURE 6 |** Cluster dendrogram based on IR characteristics in *Nidovirales*. Input data are listed in **Supplementary Material 7**. Statistically significant clusters (based on AU values above 95, equivalent to *p*-values lower than 0.05) are highlighted by rectangles drawn with broken red lines. Two main branches are highlighted by light blue and light orange backgrounds. *Nidovirales* families are marked by suffixes AR (*Arteriviridae*), COR (*Coronaviridae*), MES (*Mesnidovirineae*), MON (*Mononiviridae*), RON (*Ronidovirineae*), TOB (*Tobaniviridae*), or UN (unclassified *Nidovirales*). The green star indicates SARS-CoV and MERS and red star indicates SARS-CoV-2.



**FIGURE 7 |** Differences in IR frequency by annotated locus. The chart compares IR frequencies per 1,000 nt between “gene” annotation and other annotated locations from the NCBI database that were found in genomes more than 10 times. We analyzed frequencies of all IRs (all) and of IRs with lengths 8 nt and longer (8+), 10 nt and longer (10+), and 12 nt and longer (12+) within annotated locations (inside) as well as before and after annotated locations.

± 0.61, respectively). Wilcoxon signed-rank test with continuity correction showed the IR frequency per 1,000 nt in SARS-CoV-2 (40.23) to be significantly higher than expected (*p*-value 0.003). When we inspected differences of IR frequencies according to hosts, we found the highest mean IR frequency per 1,000 nt to be in *Nidovirales* infecting Humans (38.13), then in Vertebrates (34.25), and the lowest mean frequency is in Invertebrates (31.77).

A summary of all IRs found in ranges of different IR sizes and precomputed IR frequencies per 1,000 nt is shown in **Table 4**. Although generally the frequency of IR presence decreases with the IR length, there are notable differences between groups and also between viruses with different hosts. The most as well as longest IRs occur in the *Coronaviridae* group and in viruses having humans as a host. IRs 12 bp long and longer are very rare in the *Ronidovirineae* group. The relationship between IRs

length and stability of resulting secondary structure is not simple. While some authors believe that longer IRs are more stable, others suggest that there is an energy optimum defined by arm and spacer length (Sinden et al., 1991; Brázda et al., 2016; Georgakopoulos-Soares et al., 2018).

Differences in IR frequency according to annotated loci are shown in **Figure 5**. The most notable enrichment of IRs was found inside 5'UTR for IRs of size 12+ nt, and this is the most frequently occurring location of 12+ IRs in *Nidovirales* genomes. Noteworthy is that there are no 12+ IRs around 5'UTR loci, but there is an abundance of IRs 10+ nt long in these locations. The 5'UTR are abundant for 12+ IR, but there are no 12+ IRs within 3'UTR. This points to functional relevance of these IRs in viral genomes.

## Prediction of Human SARS-CoV-2 RNA Folding

It is very likely that coronavirus RNA in the viral capsid is compactly folded into a spiral-like structure due to nucleocapsid phosphoprotein N, as was proposed earlier for SARS-CoV (Chang et al., 2014). But what structure does the RNA take after the capsid content is released into the cytoplasm of the host cell? We made a global prediction of RNA folding for SARS-CoV-2 and in a negative control (random RNA sequences with the same GC content and length). It was clearly visible that there arose a much more compact RNA structure in SARS-CoV-2 (**Figure 8**). The RNA folding algorithm shows that inverted repeats forming hairpins have the main importance for its folding. Due to the relative lack of PQSs with high G4Hunter scores in the SARS-CoV-2 positive strand, we assume that G-quadruplexes are not important for its folding. For the 10 longest IRs obtained from Palindrome analyzer, we prepared local predictions of their 2D structures using the Mfold webserver (Zuker, 2003) (**Supplementary Material 8**).

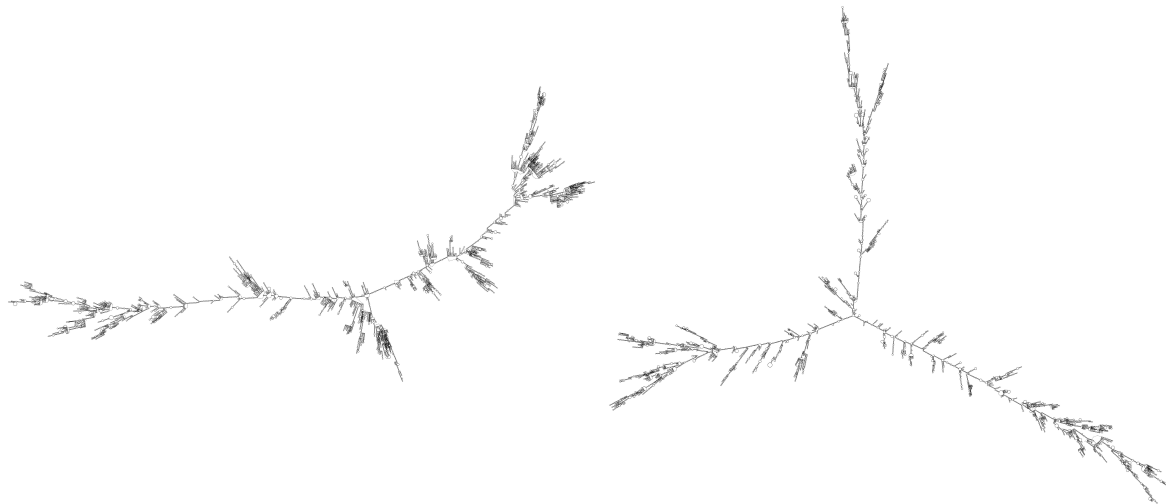
## Comparison of SUD Domain M Region of Nsp3 in Three Pathogenic Human Coronaviruses

Consisting of nearly 2,000 amino acids, Nsp3 is the largest multi-domain protein in *Nidovirales*. Nevertheless, Nsp3's role is still largely unknown. It is believed that the protein plays various roles in coronavirus infection. Nsp3 interacts with other viral Nsps as well as RNA to form the replication/transcription complex. It also acts on posttranslational modifications of host proteins to antagonize the host innate immune response. On the other hand, Nsp3 is itself modified in host cells by N-glycosylation and can interact with host proteins to support virus survival (Lei et al., 2018). Both SARS-CoVs (2003 and 2019) have PQSs in their genomes and have a retained M region of the SUD domain in protein Nsp3 that is reported to be critical for interacting with G-quadruplexes, and particularly all G-quadruplex interacting residues, as proposed by Kusov et al. (2015), are 100% conserved (**Figure 9**). MERS has no PQS in its RNA sequence and, strikingly, it has a deletion in this region of the SUD domain suggesting parallel evolution simplifications.

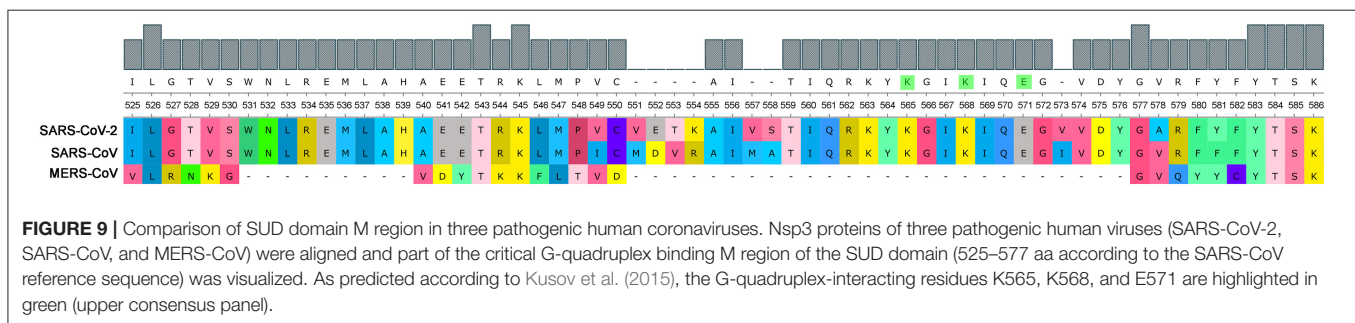
**TABLE 4 |** Total number of IRs and their resulting frequencies per 1,000 nt grouped by size of IR.

| Size of IRs                          | Number of IRs in dataset | IRs frequency per 1,000 nt |
|--------------------------------------|--------------------------|----------------------------|
| <b>All</b>                           |                          |                            |
| All                                  | 93,369                   | 33.90                      |
| 8+                                   | 12,669                   | 4.54                       |
| 10+                                  | 1,872                    | 0.65                       |
| 12+                                  | 244                      | 0.08                       |
| <b>Arteriviridae</b>                 |                          |                            |
| All                                  | 9,695                    | 29.44                      |
| 8+                                   | 1,127                    | 3.41                       |
| 10+                                  | 103                      | 0.31                       |
| 12+                                  | 8                        | 0.02                       |
| <b>Coronaviridae</b>                 |                          |                            |
| All                                  | 58,943                   | 37.08                      |
| 8+                                   | 8,527                    | 5.36                       |
| 10+                                  | 1,393                    | 0.88                       |
| 12+                                  | 190                      | 0.12                       |
| <b>Mesnidovirineae</b>               |                          |                            |
| All                                  | 8,012                    | 33.04                      |
| 8+                                   | 1,081                    | 4.46                       |
| 10+                                  | 137                      | 0.56                       |
| 12+                                  | 13                       | 0.05                       |
| <b>Ronidovirineae</b>                |                          |                            |
| All                                  | 3,474                    | 26.37                      |
| 8+                                   | 320                      | 2.44                       |
| 10+                                  | 25                       | 0.18                       |
| 12+                                  | 2                        | 0.01                       |
| <b>Tobaniviridae</b>                 |                          |                            |
| All                                  | 9,082                    | 30.24                      |
| 8+                                   | 1,060                    | 3.54                       |
| 10+                                  | 128                      | 0.43                       |
| 12+                                  | 15                       | 0.05                       |
| <b>Humans only</b>                   |                          |                            |
| All                                  | 14,408                   | 38.13                      |
| 8+                                   | 2,048                    | 5.40                       |
| 10+                                  | 344                      | 0.91                       |
| 12+                                  | 42                       | 0.11                       |
| <b>Vertebrate (Including humans)</b> |                          |                            |
| All                                  | 72,838                   | 34.25                      |
| 8+                                   | 10,097                   | 4.66                       |
| 10+                                  | 1,549                    | 0.69                       |
| 12+                                  | 207                      | 0.09                       |
| <b>Invertebrates</b>                 |                          |                            |
| All                                  | 13,233                   | 31.77                      |
| 8+                                   | 1,686                    | 4.06                       |
| 10+                                  | 218                      | 0.51                       |
| 12+                                  | 24                       | 0.05                       |
| <b>SARS-CoV-2</b>                    |                          |                            |
| All                                  | 1,203                    | 40.23                      |
| 8+                                   | 203                      | 6.79                       |
| 10+                                  | 37                       | 1.24                       |
| 12+                                  | 5                        | 0.17                       |

Shown are data for all 109 representative *Nidovirales* sequences, for families, host groups, and for SARS-CoV-2 alone. Frequency was computed using total number of IRs in each category divided by total length of all analyzed sequences and multiplied by 1,000.



**FIGURE 8 |** RNA fold (Lorenz et al., 2011) prediction for SARS-CoV-2 RNA molecule (**Left**) and random sequence negative control (**Right**). This figure shows a high level of SARS-CoV-2 genome folding via complementarity of particular RNA regions and forming of hairpins and/or cruciforms. RNA fold prediction was carried out using default parameters via Galaxy webserver (Afgan et al., 2018), which enables queries of lengths longer than 10,000 nucleotides.



**FIGURE 9 |** Comparison of SUD domain M region in three pathogenic human coronaviruses. Nsp3 proteins of three pathogenic human viruses (SARS-CoV-2, SARS-CoV, and MERS-CoV) were aligned and part of the critical G-quadruplex binding M region of the SUD domain (525–577 aa according to the SARS-CoV reference sequence) was visualized. As predicted according to Kusov et al. (2015), the G-quadruplex-interacting residues K565, K568, and E571 are highlighted in green (upper consensus panel).

## Prediction of Human RNA-Binding Proteins Sites in SARS-CoV-2 RNA

Using the RBPmap tool, we predicted human RNA-binding proteins sites in SARS-CoV-2 RNA. While holding to very stringent thresholds, three highly promising candidate human RNA-binding proteins were predicted. SRSF7 was predicted to bind viral RNA in nucleotide position 26,194–26,199 (NC\_045512.2), which is in fact exactly the binding motif for this protein (ACGACG). Protein HNRNPA1 was predicted to bind viral RNA in nucleotide position 23,090–23,097 (exact binding motif GUAGUAGU). The last protein found was TRA2A in nucleotide position 3,056–3,065 and position 3,074–3,083, in both cases with one mismatch (the motifs found were GAAGAAGAAG and the experimentally validated motif for TRA2A is GAAGAGGAAG). All three of these proteins share multiple RGG-rich novel interesting quadruplex interaction (NIQI) motifs, which are common to most human G-quadruplex-binding proteins (Brázda et al., 2018; Huang et al., 2018; Masuzawa and Oyoshi, 2020). All find individual motif occurrence (FIMO) hits below  $p = 1.10^{-4}$  are enclosed in **Supplementary Material 9**. The proteins SRSF7, HNRNPA1,

and TRA2A are involved in mRNA splicing via the spliceosome pathway (Szklarczyk et al., 2015).

## DISCUSSION

Local DNA structures have been shown to play important roles in basic biological processes, including replication and transcription (Brázda et al., 2011; Travers and Muskhelishvili, 2015; Surovtsev and Jacobs-Wagner, 2018). PQS analyses of human viruses have clearly demonstrated that these sequences are conserved also in the viral genomes (Lavezzo et al., 2018) and could be targets for antiviral therapies (Wang et al., 2015; Krafčíková et al., 2017; Ruggiero and Richter, 2018). For example, the conserved PQS sequence present in the L gene of Zaire ebolavirus and related to its replication is inhibited by interaction with G-quadruplex ligand TMPyP4, and this finding has led to suggestions that G-quadruplex RNA stabilization could constitute a new strategy against Ebola virus disease (Wang et al., 2016). A similar strategy has been proposed against Hepatitis C virus belonging to the *Flaviviridae* family of positive ssRNA viruses. Stabilization of G4s by Phen-DC3 ligand has been shown to lead to inhibited

replication of this virus (Jaubert et al., 2018). Viruses with single-stranded genomes are often non-symmetrically distributed with guanines and cytosines. For example, Zika viral genome (*Flaviviridae* family) has a G-rich positive-sense genome and a C-rich negative-sense strand. Fleming et al. found more than 60 PQSs in the Zika virus genome on the positive strand but no PQSs on the negative strand. This observation identifies a large asymmetry with respect to PQS content between the two strands. The strand asymmetry for PQS sites likely results from the high guanine content relative to cytosine content in the positive-sense strands (Fleming et al., 2016). There are many PQS searching algorithms (reviewed by Puig Lombardi and Londoño-Vallejo, 2020). We used the G4Hunter algorithm. Because it had originally been written to analyze DNA (Bedrat et al., 2016), the G4Hunter algorithm analyzed both strands simultaneously (Brázda et al., 2019). Analyses of both strands are therefore presented in our manuscript. The results with positive G4Hunter scores show PQSs in the plus strand and the results with negative G4Hunter scores show PQSs in the minus strand. In our study, we analyzed all accessible genomes of *Nidovirales*, thus including also the contemporary pandemic SARS-CoV-2 genome. Our analyses found only one PQS in the genomic sequence of SARS-CoV-2, that having a G4Hunter score of  $-1.24$  and the following sequences: 5'-CCCCAAAUCAGC GAA AUGCACCCC-3' for a positive-strand intermediate and 5'-GGGGUGCAUUUCGUGAUUUUGGGG-3' for a negative-strand intermediate (where G-quadruplex could theoretically arise). This sequence was not identified by other prediction algorithms that use searching based upon regex sequence or that are pattern-based (PQSfinder or QGRS mapper), and it has been given only a low score by the new machine-learning G4screener algorithm (Garant et al., 2018). Nevertheless, none of these algorithms consider possible substitutions of guanine by adenine in quadruplex tetrads, as has been described by Kocman and Plavec (2017), or other untypical quadruplexes as reviewed by Lightfoot et al. (2019) and stable quadruplexes with long loops that have been described (Guedin et al., 2010). Moreover, there is another GGG track in the proximity of our predicted sequence. On the other hand, analysis of the SARS-CoV-2 genome by QGRS algorithms showed 25 hits on the positive and 12 hits on the negative RNA strand (**Supplementary Material 10**). These hits are almost exclusively with two G-repeats only and have relatively low scores (maximum 19 for the QGRS algorithm and 1.111 for the G4Hunter algorithm). The best PQS suggested by the G4Hunter algorithm is located in the position 28,289–28,313 within the nucleocapsid phosphoprotein coding sequence. It is noteworthy that all viral RNAs are produced through negative-strand intermediates, which are only about 1% as abundant as their positive-sense counterparts (Fehr and Perlman, 2015). It would be of great value to know whether TMPyP4 or other G-quadruplex stabilizing compounds can inhibit replication processes of SARS-CoV-2. We have aligned three pathogenic human coronaviruses (SARS-CoV, SARS-CoV-2, and MERS-CoV) to see if there are differences in the SUD domain, which was earlier proposed to be G-quadruplex binding (Kusov et al., 2015). Comparison of three key amino acid residues involved in G-quadruplex binding revealed that the SUD domain of

MERS-CoV lacks these residues. This correlates with the fact that no G-quadruplex was predicted that is within its genome. It has been demonstrated, however, that the RGG domain can play roles in various nucleic acid and protein interactions (Thandapani et al., 2013), so this correlation can be G4-independent for SARS-CoV-2. RNA hairpins, which are formed by IRs, are basic structural elements of RNA and play crucial roles in gene expression and intermolecular recognition. Conserved palindromic RNA structures have been found in many viral genomes, including HIV-1, and play a crucial role in their replication (Liu et al., 2018). We have found an abundance of IRs inside 5'UTR in *Nidovirales* genomes. In general, 5'UTR is an important locus for regulation of viral replication and gene expression. It has been demonstrated that stem integrity of phylogenetically conserved stem-loop structure located in 5'UTR of the PRRSV virus from the *Arteriviridae* family is crucial for viral replication and subgenomic mRNA synthesis. Similar secondary structures have been proposed for several viruses from the *Arteriviridae* and *Coronaviridae* families (Lu et al., 2011). Our analyses of annotated features in **Figure 5** therefore support this report. The discrete locations of specific IRs in viral genomes could therefore be additional targets for their regulation.

Significant differences of PQS and IR frequencies among various ssRNA viruses in *Nidovirales* groups show that their genome organization and regulation are not identical but that for some *Nidovirales* the presence of the G-quadruplexes most probably does not play an essential role in their biological regulation. Moreover, our analyses suggest that G-quadruplexes have been evolutionarily eliminated in some genomes of *Nidovirales*. This is quite surprising, considering that G-quadruplexes have been found in all evolutionary groups, including such CG low-level organisms as *Saccharomyces cerevisiae* (Bartas et al., 2019; Brázda et al., 2019; Singh and Lakhnypaul, 2019; Gage and Merrick, 2020). In fact, it could be an evolutionary advantage not to present G-quadruplex in viral genome because a number of cellular proteins interact with G-quadruplexes (Brázda et al., 2014; Mishra et al., 2016) and therefore the presence of G-quadruplex in viral genome could serve as a structure recognized by the innate immune system (Unterholzner et al., 2010; Hároníková et al., 2016; Voter et al., 2020). Their RNAs are therefore not recognized as alien nucleic acids and are processed by the cellular machinery. On the other hand, presence of IRs in *Nidovirales* genomes constitutes an inseparable part of their genomes and allows their correct folding and structure-specific regulation of their functions (Lorenz et al., 2011; Dutkiewicz et al., 2016).

Targeting viral proteins is usually effective only against specific viral strains and fails even for closely related viral species. It appears that targeting host proteins should be able to provide a response toward a wider spectrum of viruses inasmuch as different viruses exploit common cellular pathways. Many cellular RNA-binding proteins (RBPs) containing well-established RNA-binding domains (RBDs) are known to be critical for infection by different viruses. Recently, 472 RBPs were reviewed for their linkage to viruses (García-Moreno et al., 2018). It has been demonstrated that G-quadruplex formation in HIV-1 viral genome stalls RNA polymerase, thereby limiting

viral replication in host cells. HIV-1 nucleocapsid protein NCp7 helps to resolve G-quadruplex formation and therefore enables virus to spread. Stabilization of this quadruplex has been targeted by several experimental compounds. This treatment slowed or inhibited viral growth (Butovskaya et al., 2019).

Virus reproduction is dependent on the cellular transcription machinery, and therefore the interaction of the cellular proteins with viral RNA could be another target for antiviral therapy (Roberts et al., 2009). Our analysis predicted several QBPs to be capable to bind SARS-CoV-2 RNA. One of these, HNRNPA1 protein, is responsible for nuclear-cytoplasm shuttling (Garcia-Moreno et al., 2018). Like some other RNA-binding proteins, HNRNPA1 forms so-called membraneless organelles. These organelles are assemblies of proteins along with RNA or DNA that condense in specific cellular loci. The organelles can undergo transition from liquid-like droplets to amyloid fibrils, and mutations in so-called low complexity domains of these proteins lead to formation of amyloid aggregations that are found in many neurodegenerative diseases (Gui et al., 2019). HNRNPA1 is involved in many different cellular processes and has been targeted in various diseases. One such example is the varicella zoster virus that causes chicken pox. Moreover, this response to varicella zoster virus has been connected with autoimmune disease that complicates multiple sclerosis (Kattimani and Veerappa, 2018). HNRNPA1 is known to bind and resolve G-quadruplex formed in TRA2B promoter and to promote its transcription. Dysregulation of this binding leads to progression of colon cancer. This interaction has been targeted by the well-known G-quadruplex stabilizer pyridostatin, which led to decreased transcription from the TRA2B promoter (Nishikawa et al., 2019). Furthermore, HNRNPA1 is co-expressed with bromodomain and extraterminal domain protein BRD4 in human tumor samples. It has been shown experimentally that the well-described, naturally occurring polyphenolic flavonoid quercetin inhibited this protein and thereby led to better susceptibility to treatment in cancer patients (Pham et al., 2019). Both SRSF7 and TRA2A proteins, which are predicted to interact with SARS-CoV-2 RNA, play roles in alternative splicing (Ghosh et al., 2016). SRSF7 is a serine and arginine-rich splicing factor and is part of the spliceosome. Its expression has been connected to several types of cancer and it has been shown that its knockdown induced p21 expression and thus reduced cancer development (Saijo et al., 2016). Little is known about TRA2A protein. It was first identified in insects together with its paralog TRA2B, which has been studied to a greater extent (Tan et al., 2018). It has been found that TRA2A can promote paclitaxel therapy and promote cancer progression in triple-negative breast cancers (Liu et al., 2017). A role of TRA2A protein in regulation of HIV1 virus replication has been described. Both TRA2A and TRA2B bind to a specific HIV1 sequence and regulate its replication within the cell through alternative splicing of viral RNA (Erkelenz et al., 2013).

Scientists around the world are united in their efforts to find an effective therapy against coronavirus disease (COVID-19). Among the most promising candidates are remdesivir and chloroquine. Remdesivir is an adenosine analog that incorporates into nascent viral RNA chains and results in premature

termination. Preliminary data have shown that remdesivir effectively inhibited virus infection in a human liver cancer cell line (Wang et al., 2020). Chloroquine is a potential broad-spectrum antiviral drug and it already has been widely used as a low-cost and safe anti-malarial and autoimmune disease drug for more than 70 years. Application of chloroquine causes elevation of endosomal pH and also interferes with terminal glycosylation of the cellular receptor ACE2. This probably has a negative influence on virus-receptor binding and abrogates the infection (Vincent et al., 2005). Drug repurposing seems like a very good strategy for quickly and at low cost finding a new therapy for new human diseases (Oprea and Mestres, 2012). To date, there are few substances with G-quadruplex stabilizing features. One example is topotecan (also known by its brand name Hycamtin®), which frequently is used for treating ovary cancer. If everything else fails, it seems to be an option. On the other hand, topotecan cause unpleasant side effects, such as nausea, vomiting, and diarrhea (Topotecan - Chemotherapy Drugs - Chemocare, n.d.). Many studies have described strong G-quadruplex stabilization effects (Li et al., 2018; Satpathi et al., 2018), which might be one possible mode of action. Moreover, it has been proven that higher structures of nucleic acids, and G-quadruplex especially, might be stabilized by use of various natural substances. Berbamine is one such substance and is a component of traditional Chinese medicine. It frequently is used for treating chronic myeloid leukemia or melanoma and has strong binding affinity to G-quadruplex structures (Tan et al., 2014). Viral nucleic acids and their loci with G-quadruplex-forming potential are in all cases very specific and are promising molecular targets for treating serious diseases. Evidence of G-quadruplex formation as a potential target for therapy was proven for Hepatitis A, flu virus, and HIV-1 (Métifiot et al., 2014). Because all the aforementioned cases concern RNA viruses, the viral G-quadruplexes are localized in the cytoplasm of the host cell. Our comparative analyses of IRs and PQSs in *Nidovirales* show that, in contrast to IRs, which are presented abundantly in all *Nidovirales* genomes, the sequences able to form G-quadruplex structures are very unequally distributed and are very rare especially in *Nidovirales* species capable to infect humans. This suggests intentional suppression during evolution in order to simplify viral RNA replication. Finding the proper stabilizers of viral higher RNA structures might be crucial for inhibiting or stopping viral RNA replication in order to gain time for the immune system to deal successfully with an infection.

## DATA AVAILABILITY STATEMENT

All datasets presented in this study are included in the article/**Supplementary Material**.

## AUTHOR CONTRIBUTIONS

MB and VB contributed to the conceptualization, formal analysis, and methodology. NB, AC, AV, and TS helped with the data curation. PP was involved in with funding acquisition

and project administration. MB, VB, AV, and JČ carried out the investigation. MB, NB, OP, VB, and EJ helped with the resources. VB, KM, and PP supervised the study. VB and EJ validated the study. MB, NB, and AC worked on the visualization. MB, VB, NB, OP, and JČ wrote the original draft. VB, JČ, KM, and PP reviewed and edited the manuscript. All authors contributed to the article and approved the submitted version.

## FUNDING

This work was supported by the Ministry of Education, Youth and Sports of the Czech Republic in the National Feasibility Program I (LO1208 TEWEP); by the EU structural funding Operational Programme Research and Development

## REFERENCES

- Afgan, E., Baker, D., Batut, B., Van Den Beek, M., Bouvier, D., Cech, M., et al. (2018). The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. *Nucl. Acids Res.* 46, W537–W544. doi: 10.1093/nar/gky379
- Bagga, R., Ramesh, N., and Brahmachari, S. K. (1990). Supercoil-induced unusual DNA structures as transcriptional block. *Nucl. Acids Res.* 18, 3363–3369. doi: 10.1093/nar/18.11.3363
- Bartas, M., Cutová, M., Brázda, V., Kaura, P., Štátný, J., Kolomazník, J., et al. (2019). The presence and localization of G-quadruplex forming sequences in the domain of bacteria. *Molecules* 24:1711. doi: 10.3390/molecules24091711
- Bedrat, A., Lacroix, L., and Mergny, J.-L. (2016). Re-evaluation of G-quadruplex propensity with G4Hunter. *Nucl. Acids Res.* 44, 1746–1759. doi: 10.1093/nar/gkw006
- Brázda, V., Cerven, J., Bartas, M., Mikysková, N., Coufal, J., Pečinka, P., et al. (2018). The amino acid composition of quadruplex binding proteins reveals a shared motif and predicts new potential quadruplex interactors. *Molecules* 23:2341. doi: 10.3390/molecules23092341
- Brázda, V., Hároníková, L., Liao, J. C., and Fojta, M. (2014). DNA and RNA quadruplex-binding proteins. *Int. J. Mol. Sci.* 15, 17493–17517. doi: 10.3390/ijms151017493
- Brázda, V., Kolomazník, J., Lýsek, J., Bartas, M., Fojta, M., Štátný, J., et al. (2019). G4Hunter web application: a web server for G-quadruplex prediction. *Bioinformatics* 35, 3493–3495. doi: 10.1093/bioinformatics/btz087
- Brázda, V., Kolomazník, J., Lýsek, J., Hároníková, L., Coufal, J., and Štátný, J. (2016). Palindrome analyser—A new web-based server for predicting and evaluating inverted repeats in nucleotide sequences. *Biochem. Biophys. Res. Commun.* 478, 1739–1745. doi: 10.1016/j.bbrc.2016.09.015
- Brázda, V., Laister, R. C., Jagelská, E. B., and Arrowsmith, C. (2011). Cruciform structures are a common DNA feature important for regulating biological processes. *BMC Mol. Biol.* 12:33. doi: 10.1186/1471-2199-12-33
- Bridges, R., Correia, S., Wegner, F., Venturini, C., Palsler, A., White, R. E., et al. (2019). Essential role of inverted repeat in Epstein-Barr virus IR-1 in B cell transformation; geographical variation of the viral genome. *Philos. Trans. R. Soc. B* 374:20180299. doi: 10.1098/rstb.2018.0299
- Burge, S., Parkinson, G. N., Hazel, P., Todd, A. K., and Neidle, S. (2006). Quadruplex DNA: sequence, topology and structure. *Nucl. Acids Res.* 34, 5402–5415. doi: 10.1093/nar/gkl655
- Butovskaya, E., Solda, P., Scalabrini, M., Nadai, M., and Richter, S. N. (2019). HIV-1 nucleocapsid protein unfolds stable RNA G-quadruplexes in the viral genome and is inhibited by G-quadruplex ligands. *ACS Infect. Dis.* 5, 2127–2135. doi: 10.1021/acsinfecdis.9b00272
- Carrasco-Hernandez, R., Jácome, R., López Vidal, Y., and Ponce de León, S. (2017). Are RNA viruses candidate agents for the next global pandemic? A review. *ILAR J.* 58, 343–358. doi: 10.1093/ilar/ilx026
- Chambers, V. S., Marsico, G., Boutell, J. M., Di Antonio, M., Smith, G. P., and Balasubramanian, S. (2015). High-throughput sequencing of DNA G-quadruplex structures in the human genome. *Nat. Biotechnol.* 33:877. doi: 10.1038/nbt.3295
- Chang, C., Hou, M.-H., Chang, C.-F., Hsiao, C.-D., and Huang, T. (2014). The SARS coronavirus nucleocapsid protein – Forms and functions. *Antiviral Res.* 103, 39–50. doi: 10.1016/j.antiviral.2013.12.009
- Chen, N., Li, X., Li, S., Xiao, Y., Ye, M., Yan, X., et al. (2020). How related is SARS-CoV-2 to other coronaviruses? *Vet. Rec.* 186, 496–496. doi: 10.1136/vr.m1452
- Cohen, J. (2020). New coronavirus threat galvanizes scientists. *Science* 367, 492–493. doi: 10.1126/science.367.6477.492
- Cowling, B. J., Park, M., Fang, V. J., Wu, P., Leung, G. M., and Wu, J. T. (2015). Preliminary epidemiologic assessment of MERS-CoV outbreak in South Korea, May–June 2015. *Euro Surveill.* 20:21163. doi: 10.2807/1560-7917.ES2015.20.25.21163
- Drake, J. W., and Holland, J. J. (1999). Mutation rates among RNA viruses. *Proc. Natl. Acad. Sci. U. S. A.* 96, 13910–13913. doi: 10.1073/pnas.96.24.13910
- Dutkiewicz, M., Stachowiak, A., Swiatkowska, A., and Ciesiolka, J. (2016). Structure and function of RNA elements present in enteroviral genomes. *Acta Biochim. Polonica* 63, 623–630. doi: 10.18388/abp.2016\_1337
- Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucl. Acids Res.* 32, 1792–1797. doi: 10.1093/nar/gkh340
- Eigen, M., and Schuster, P. (1977). A principle of natural self-organization: Part A: emergence of the hypercycle. *Naturwissenschaften* 64, 541–565. doi: 10.1007/BF00450633
- Erkelenz, S., Poschmann, G., Theiss, S., Stefanski, A., Hillebrand, F., Otte, M., et al. (2013). Tra2-mediated recognition of HIV-1 5' splice site D3 as a key factor in the processing of vpr mRNA. *J. Virol.* 87, 2721–2734. doi: 10.1128/JVI.02756-12
- Federhen, S. (2011). The NCBI taxonomy database. *Nucl. Acids Res.* 40, D136–D143. doi: 10.1093/nar/gkr1178
- Fehr, A. R., and 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
- Fleming, A. M., Ding, Y., Alenko, A., and Burrows, C. J. (2016). Zika virus genomic RNA possesses conserved G-quadruplexes characteristic of the flaviviridae family. *ACS Infect. Dis.* 2, 674–681. doi: 10.1021/acsinfecdis.6b00109
- Frasson, I., Nadai, M., and Richter, S. N. (2019). Conserved G-quadruplexes regulate the immediate early promoters of human Alphaherpesviruses. *Molecules* 24:2375. doi: 10.3390/molecules24132375
- Gage, H. L., and Merrick, C. J. (2020). Conserved associations between G-quadruplex-forming DNA motifs and virulence gene families in malaria parasites. *BMC Genomics* 21:236. doi: 10.1186/s12864-020-6625-x
- Garant, J.-M., Perreault, J.-P., and Scott, M. S. (2018). G4RNA screener web server: user focused interface for RNA G-quadruplex prediction. *Biochimie* 151, 115–118. doi: 10.1016/j.biochi.2018.06.002

for innovation, project No. CZ.1.05/2.1.00/19.0388; and by the projects SGS/01/PrF/2020 and SGS/07/PrF/2020 financed by University of Ostrava.

## ACKNOWLEDGMENTS

We would like to express our gratitude to Alena Volná, M.Sc. for her time spent in preparing the draft illustrations accompanying this work. **Figure 1** is reproduced with permission from Biolution GmbH.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmicb.2020.01583/full#supplementary-material>

- Garcia-Moreno, M., Jaervelin, A. I., and Castello, A. (2018). Unconventional RNA-binding proteins step into the virus-host battlefield. *Wiley Interdiscipl. Rev. RNA* 9:e1498. doi: 10.1002/wrna.1498
- Georgakopoulos-Soares, I., Morganello, S., Jain, N., Hemberg, M., and Nik-Zainal, S. (2018). Noncanonical secondary structures arising from non-B DNA motifs are determinants of mutagenesis. *Genome Res.* 28, 1264–1271. doi: 10.1101/gr.231688.117
- Ghosh, P., Grellscheid, S. N., and Sowdhamini, R. (2016). A tale of two paralogs: human Transformer2 proteins with differential RNA-binding affinities. *J. Biomol. Struct. Dyn.* 34, 1979–1986. doi: 10.1080/07391102.2015.1100551
- Gorbalenya, A. E. (2001). Big nidovirus genome. When count and order of domains matter. *Adv. Exp. Med. Biol.* 494, 1–17. doi: 10.1007/978-1-4615-1325-4\_1
- Gorbalenya, A. E., Enjuanes, L., Ziebuhr, J., and Snijder, E. J. (2006). Nidovirales: evolving the largest RNA virus genome. *Virus Res.* 117, 17–37. doi: 10.1016/j.virusres.2006.01.017
- Guedin, A., Gros, J., Alberti, P., and Mergny, J.-L. (2010). How long is too long? Effects of loop size on G-quadruplex stability. *Nucl. Acids Res.* 38, 7858–7868. doi: 10.1093/nar/gkq639
- Gui, X., Luo, F., Li, Y., Zhou, H., Qin, Z., Liu, Z., et al. (2019). Structural basis for reversible amyloids of hnRNP1 elucidates their role in stress granule assembly. *Nat. Commun.* 10:2006. doi: 10.1038/s41467-019-09902-7
- Hanada, K., Suzuki, Y., Nakane, T., Hirose, O., and Gojobori, T. (2005). The origin and evolution of porcine reproductive and respiratory syndrome viruses. *Mol. Biol. Evol.* 22, 1024–1031. doi: 10.1093/molbev/msi089
- Hänsel-Hertsch, R., Beraldi, D., Lensing, S. V., Marsico, G., Zyner, K., Parry, A., et al. (2016). G-quadruplex structures mark human regulatory chromatin. *Nat. Genet.* 48:1267. doi: 10.1038/ng.3662
- Hároníková, L., Coufal, J., Kejnovská, I., Jagelská, E. B., Fojta, M., Dvoráková, P., et al. (2016). IFI16 preferentially binds to DNA with quadruplex structure and enhances DNA quadruplex formation. *PLoS ONE* 11:e0157156. doi: 10.1371/journal.pone.0157156
- Hoffmann, M., Kleine-Weber, H., Schroeder, S., Krüger, N., Herrler, T., Erichsen, S., et al. (2020). SARS-CoV-2 cell entry depends on ACE2 and TMPRSS2 and is blocked by a clinically proven protease inhibitor. *Cell.* 181, 271–280.e8. doi: 10.1016/j.cell.2020.02.052
- Huang, Z.-L., Dai, J., Luo, W.-H., Wang, X.-G., Tan, J.-H., Chen, S.-B., et al. (2018). Identification of G-quadruplex-binding protein from the exploration of RGG Motif/G-quadruplex interactions. *J. Am. Chem. Soc.* 140, 17945–17955. doi: 10.1021/jacs.8b09329
- Hung, L. S. (2003). The SARS epidemic in Hong Kong: what lessons have we learned? *J. R. Soc. Med.* 96, 374–378. doi: 10.1258/jrsm.96.8.374
- Ishimaru, D., Plant, E. P., Sims, A. C., Yount, B. L. Jr., Roth, B. M., Eldho, N. V., et al. (2013). RNA dimerization plays a role in ribosomal frameshifting of the SARS coronavirus. *Nucl. Acids Res.* 41, 2594–2608. doi: 10.1093/nar/gks1361
- Jaubert, C., Bedrat, A., Bartolucci, L., Di Primo, C., Ventura, M., Mergny, J.-L., et al. (2018). RNA synthesis is modulated by G-quadruplex formation in Hepatitis C virus negative RNA strand. *Sci. Rep.* 8, 1–9. doi: 10.1038/s41598-018-26582-3
- Kattimani, Y., and Veerappa, A. M. (2018). Complex interaction between mutant HNRNP1 and gE of varicella zoster virus in pathogenesis of multiple sclerosis. *Autoimmunity* 51, 147–151. doi: 10.1080/08916934.2018.1482883
- Kikin, O., D'Antonio, L., and Bagga, P. S. (2006). QGRS Mapper: a web-based server for predicting G-quadruplexes in nucleotide sequences. *Nucl. Acids Res.* 34, W676–W682. doi: 10.1093/nar/gkl253
- Kim, D., Lee, J.-Y., Yang, J.-S., Kim, J. W., Kim, V. N., and Chang, H. (2020). The architecture of SARS-CoV-2 transcriptome. *bioRxiv.* 181, 914–921.e10. doi: 10.1016/j.cell.2020.04.011
- Kocman, V., and Plavec, J. (2017). Tetrahelical structural family adopted by AGCGA-rich regulatory DNA regions. *Nat. Commun.* 8:15355. doi: 10.1038/ncomms15355
- Kolesnikova, S., and Curtis, E. A. (2019). Structure and function of multimeric G-quadruplexes. *Molecules* 24:3074. doi: 10.3390/molecules24173074
- Krafcíková, P., Demkovičová, E., and Víglašký, V. (2017). Ebola virus derived G-quadruplexes: thiazole orange interaction. *Biochim. Biophys. Acta General Subj.* 1861, 1321–1328. doi: 10.1016/j.bbagen.2016.12.009
- Kusov, Y., Tan, J., Alvarez, E., Enjuanes, L., and Hilgenfeld, R. (2015). A G-quadruplex-binding macrodomain within the “SARS-unique domain” is essential for the activity of the SARS-coronavirus replication–transcription complex. *Virology* 484, 313–322. doi: 10.1016/j.virol.2015.06.016
- Lai, M. M. C., and Cavanagh, D. (1997). “The molecular biology of coronaviruses,” in *Advances in Virus Research* (Elsevier), 1–100. Available online at: <https://linkinghub.elsevier.com/retrieve/pii/S0065352708602869>
- Lauring, A. S., Frydman, J., and Andino, R. (2013). The role of mutational robustness in RNA virus evolution. *Nat. Rev. Microbiol.* 11, 327–336. doi: 10.1038/nrmicro3003
- Lavezzo, E., Berselli, M., Frasson, I., Perrone, R., Palù, G., Brazzale, A. R., et al. (2018). G-quadruplex forming sequences in the genome of all known human viruses: a comprehensive guide. *PLoS Comput. Biol.* 14:e1006675. doi: 10.1371/journal.pcbi.1006675
- Lei, J., Kusov, Y., and Hilgenfeld, R. (2018). Nsp3 of coronaviruses: Structures and functions of a large multi-domain protein. *Antiviral Res.* 149, 58–74. doi: 10.1016/j.antiviral.2017.11.001
- Letunic, I., and Bork, P. (2019). Interactive Tree Of Life (iTOL) v4: recent updates and new developments. *Nucl. Acids Res.* 47, W256–W259. doi: 10.1093/nar/gkz239
- Li, F., Zhou, J., Xu, M., and Yuan, G. (2018). Exploration of G-quadruplex function in c-Myb gene and its transcriptional regulation by topotecan. *Int. J. Biol. Macromol.* 107, 1474–1479. doi: 10.1016/j.ijbiomac.2017.10.010
- Li, Q., Guan, X., Wu, P., Wang, X., Zhou, L., Tong, Y., et al. (2020). Early transmission dynamics in Wuhan, China, of novel coronavirus-infected pneumonia. *N. Engl. J. Med.* 382, 1199–1207. doi: 10.1056/NEJMoa2001316
- Li, R.-F., and Li, H. (2010). Study on the influences of palindromes in protein coding sequences on the folding rates of peptide chains. *Protein Peptide Lett.* 17, 881–888. doi: 10.2174/092986610791306652
- Lightfoot, H. L., Hagen, T., Tatum, N. J., and Hall, J. (2019). The diverse structural landscape of quadruplexes. *FEBS Lett.* 593, 2083–2102. doi: 10.1002/1873-3468.13547
- Liò, P., and Goldman, N. (2004). Phylogenomics and bioinformatics of SARS-CoV. *Trends Microbiol.* 12, 106–111. doi: 10.1016/j.tim.2004.01.005
- Liu, T., Sun, H., Zhu, D., Dong, X., Liu, F., Liang, X., et al. (2017). TRA2A promoted paclitaxel resistance and tumor progression in triple-negative breast cancers via regulating alternative splicing. *Mol. Cancer Therap.* 16, 1377–1388. doi: 10.1158/1535-7163.MCT-17-0026
- Liu, Y., Chen, J., Nikolaichik, O. A., Desimie, B. A., Busan, S., Pathak, V. K., et al. (2018). The roles of five conserved lentiviral RNA structures in HIV-1 replication. *Virology* 514, 1–8. doi: 10.1016/j.virol.2017.10.020
- Lorenz, R., Bernhart, S. H., Höner zu Siederdissen, C., Tafer, H., Flamm, C., Stadler, P. F., et al. (2011). ViennaRNA package 2.0. *Algorith. Mol. Biol.* 6:26. doi: 10.1186/1748-7188-6-26
- Lu, J., Gao, F., Wei, Z., Liu, P., Liu, C., Zheng, H., et al. (2011). A 5′-proximal stem-loop structure of 5′ untranslated region of porcine reproductive and respiratory syndrome virus genome is key for virus replication. *Viol. J.* 8:172. doi: 10.1186/1743-422X-8-172
- Lu, R., Zhao, X., Li, J., Niu, P., Yang, B., Wu, H., et al. (2020). Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. *Lancet* 395, 565–574. doi: 10.1016/S0140-6736(20)30251-8
- Masuzawa, T., and Oyoshi, T. (2020). Roles of the RGG domain and RNA recognition motif of nucleolin in G-quadruplex stabilization. *ACS Omega.* 5, 5202–5208. doi: 10.1021/acsomega.9b04221
- Métifiot, M., Amrane, S., Litvak, S., and Andreola, M.-L. (2014). G-quadruplexes in viruses: function and potential therapeutic applications. *Nucl. Acids Res.* 42, 12352–12366. doi: 10.1093/nar/gku999
- Mishra, S. K., Tawani, A., Mishra, A., and Kumar, A. (2016). G4IPDB: A database for G-quadruplex structure forming nucleic acid interacting proteins. *Sci. Rep.* 6:38144. doi: 10.1038/srep38144
- Modrow, S., Falke, D., Truyen, U., and Schätzl, H. (2013). “Viruses with single-stranded, positive-sense RNA genomes,” in *Molecular Virology* (Berlin: Springer Berlin Heidelberg), 185–349. Available online at: [http://link.springer.com/10.1007/978-3-642-20718-1\\_14](http://link.springer.com/10.1007/978-3-642-20718-1_14)
- Nishikawa, T., Kuwano, Y., Takahara, Y., Nishida, K., and Rokutan, K. (2019). HnRNP1 interacts with G-quadruplex in the TRA2B promoter and stimulates its transcription in human colon cancer cells. *Sci. Rep.* 9:10276. doi: 10.1038/s41598-019-46659-x

- Oboho, I. K., Tomczyk, S. M., Al-Asmari, A. M., Banjar, A. A., Al-Mugti, H., Aloraini, M. S., et al. (2015). 2014 MERS-CoV outbreak in Jeddah—a link to health care facilities. *N. Engl. J. Med.* 372, 846–854. doi: 10.1056/NEJMoa1408636
- Okonechnikov, K., Golosova, O., Fursov, M., and Team, U. (2012). Unipro UGENE: a unified bioinformatics toolkit. *Bioinformatics* 28, 1166–1167. doi: 10.1093/bioinformatics/bts091
- Oprea, T. I., and Mestres, J. (2012). Drug repurposing: far beyond new targets for old drugs. *AAPS J.* 14, 759–763. doi: 10.1208/s12248-012-9390-1
- Patino-Galindo, J. A., Filip, I., AlQuraishi, M., and Rabadan, R. (2020). Recombination and lineage-specific mutations led to the emergence of SARS-CoV-2. *bioRxiv [Preprint]*. doi: 10.1101/2020.02.10.942748
- Paz, I., Kosti, I., Ares, M. Jr., Cline, M., and Mandel-Gutfreund, Y. (2014). RBPmap: a web server for mapping binding sites of RNA-binding proteins. *Nucl. Acids Res.* 42, W361–W367. doi: 10.1093/nar/gku406
- Pham, T. N. D., Stempel, S., Shields, M. A., Spaulding, C., Kumar, K., Bentrem, D. J., et al. (2019). Quercetin enhances the anti-tumor effects of BET inhibitors by suppressing hnRNPA1. *Int. J. Mol. Sci.* 20:4293. doi: 10.3390/ijms20174293
- Plant, E., Perez-Alvarado, G., Jacobs, J., Mukhopadhyay, B., Hennig, M., and Dinman, J. (2005). A three-stemmed mRNA pseudoknot in the SARS coronavirus frameshift signal. *PLoS Biol.* 3, 1012–1023. doi: 10.1371/journal.pbio.0030172
- Platella, C., Riccardi, C., Montesarchio, D., Roviello, G. N., and Musumeci, D. (2017). G-quadruplex-based aptamers against protein targets in therapy and diagnostics. *Biochim. Biophys. Acta General Subj.* 1861, 1429–1447. doi: 10.1016/j.bbagen.2016.11.027
- Puig Lombardi, E., and Londoño-Vallejo, A. (2020). A guide to computational methods for G-quadruplex prediction. *Nucl. Acids Res.* 48, 1–15. doi: 10.1093/nar/gkaa033
- Roberts, L. O., Jopling, C. L., Jackson, R. J., and Willis, A. E. (2009). “Chapter 9 viral strategies to subvert the mammalian translation machinery,” in *Progress in Molecular Biology and Translational Science, Translational Control in Health and Disease*. (Academic Press), 313–367. Available online at: <http://www.sciencedirect.com/science/article/pii/S1877117309900096>
- Ruggiero, E., and Richter, S. N. (2018). G-quadruplexes and G-quadruplex ligands: targets and tools in antiviral therapy. *Nucl. Acids Res.* 46, 3270–3283. doi: 10.1093/nar/gky187
- Ruggiero, E., and Richter, S. N. (2020). Viral G-quadruplexes: new frontiers in virus pathogenesis and antiviral therapy. *Annu. Rep. Med. Chem.* doi: 10.1016/bs.armc.2020.04.001. [Epub ahead of print].
- Saberi, A., Gulyaeva, A. A., Brubacher, J. L., Newmark, P. A., and Gorbalenya, A. E. (2018). A planarian nidovirus expands the limits of RNA genome size. *PLoS Pathog.* 14:e1007314. doi: 10.1371/journal.ppat.1007314
- Saijo, S., Kuwano, Y., Masuda, K., Nishikawa, T., Rokutan, K., and Nishida, K. (2016). Serine/arginine-rich splicing factor 7 regulates p21-dependent growth arrest in colon cancer cells. *J. Med. Invest.* 63, 219–226. doi: 10.2152/jmi.63.219
- Satpathi, S., Singh, R. K., Mukherjee, A., and Hazra, P. (2018). Controlling anticancer drug mediated G-quadruplex formation and stabilization by a molecular container. *Phys. Chem. Chem. Phys.* 20, 7808–7818. doi: 10.1039/C8CP00325D
- Sinden, R., Zheng, G., Brankamp, R., and Allen, K. (1991). On the deletion of inverted repeated DNA in *Escherichia coli*: effects of length, thermal stability, and cruciform formation in vivo. *Genetics* 129, 991–1005.
- Singh, A., and Lakhanpaul, S. (2019). Genome-wide analysis of putative G-quadruplex sequences (PGQs) in onion yellows phytoplasma (strain OY-M): an emerging plant pathogenic bacteria. *Indian J. Microbiol.* 59, 468–475. doi: 10.1007/s12088-019-00831-z
- Snijder, E. J., Bredenbeek, P. J., Dobbe, J. C., Thiel, V., Ziebuhr, J., Poon, L. L. M., et al. (2003). Unique and conserved features of genome and proteome of SARS-coronavirus, an early split-off from the coronavirus group 2 lineage. *J. Mol. Biol.* 331, 991–1004. doi: 10.1016/S0022-2836(03)00865-9
- Surovtsev, I. V., and Jacobs-Wagner, C. (2018). Subcellular organization: a critical feature of bacterial cell replication. *Cell* 172, 1271–1293. doi: 10.1016/j.cell.2018.01.014
- Szklarczyk, D., Franceschini, A., Wyder, S., Forslund, K., Heller, D., Huerta-Cepas, J., et al. (2015). STRING v10: protein-protein interaction networks, integrated over the tree of life. *Nucl. Acids Res.* 43, D447–52. doi: 10.1093/nar/gku1003
- Tan, W., Zhou, J., and Yuan, G. (2014). Electrospray ionization mass spectrometry probing of binding affinity of berbamine, a flexible cyclic alkaloid from traditional Chinese medicine, with G-quadruplex DNA. *Rapid Commun. Mass Spectr.* 28, 143–147. doi: 10.1002/rcm.6763
- Tan, Y., Hu, X., Deng, Y., Yuan, P., Xie, Y., and Wang, J. (2018). TRA2A promotes proliferation, migration, invasion and epithelial mesenchymal transition of glioma cells. *Brain Res. Bull.* 143, 138–144. doi: 10.1016/j.brainresbull.2018.10.006
- Thandapani, P., O'Connor, T. R., Bailey, T. L., and Richard, S. (2013). Defining the RGG/RG motif. *Mol. Cell* 50, 613–623. doi: 10.1016/j.molcel.2013.05.021
- Topotecan - Chemotherapy Drugs - Chemocare. Available online at: <http://chemocare.com/chemotherapy/drug-info/Topotecan.aspx>
- Travers, A., and Muskhelishvili, G. (2015). DNA structure and function. *FEBS J.* 2279–2295. doi: 10.1111/febs.13307
- Unterholzner, L., Keating, S. E., Baran, M., Horan, K. A., Jensen, S. B., Sharma, S., et al. (2010). IFI16 is an innate immune sensor for intracellular DNA. *Nat. Immunol.* 11, 997–1004. doi: 10.1038/ni.1932
- Vincent, M. J., Bergeron, E., Benjannet, S., Erickson, B. R., Rollin, P. E., Ksiazek, T. G., et al. (2005). Chloroquine is a potent inhibitor of SARS coronavirus infection and spread. *Viol. J.* 2:69. doi: 10.1186/1743-422X-2-69
- Vorlíčková, M., Kejnovská, I., Bednářová, K., Renčíuk, D., and Kypr, J. (2012). Circular dichroism spectroscopy of DNA: from duplexes to quadruplexes. *Chirality* 24, 691–698. doi: 10.1002/chir.22064
- Voter, A. F., Callaghan, M. M., Tippana, R., Myong, S., Dillard, J. P., and Keck, J. L. (2020). Antigenic variation in neisseria gonorrhoeae occurs independently of RecQ-mediated unwinding of the pilE G quadruplex. *J. Bacteriol.* 202:e00607-19. doi: 10.1128/JB.00607-19
- Wang, M., Cao, R., Zhang, L., Yang, X., Liu, J., Xu, M., et al. (2020). Remdesivir and chloroquine effectively inhibit the recently emerged novel coronavirus (2019-nCoV) in vitro. *Cell Res.* 30, 269–271. doi: 10.1038/s41422-020-0282-0
- Wang, S.-K., Wu, Y., and Ou, T.-M. (2015). RNA G-quadruplex: the new potential targets for therapy. *Curr. Topics Med. Chem.* 15, 1947–1956. doi: 10.2174/1568026615666150515145733
- Wang, S.-R., Zhang, Q.-Y., Wang, J.-Q., Ge, X.-Y., Song, Y.-Y., Wang, Y.-F., et al. (2016). Chemical targeting of a G-quadruplex RNA in the Ebola virus L gene. *Cell Chem. Biol.* 23, 1113–1122. doi: 10.1016/j.chembiol.2016.07.019
- Wu, F., Zhao, S., Yu, B., Chen, Y.-M., Wang, W., Song, Z.-G., et al. (2020). A new coronavirus associated with human respiratory disease in China. *Nature* 579, 265–269. doi: 10.1038/s41586-020-2008-3
- Wu, Y., Shin-ya, K., and Brosh, R. M. Jr. (2008). FANCD1 helicase defective in Fanconi anemia and breast cancer unwinds G-quadruplex DNA to defend genomic stability. *Mol. Cell Biol.* 28, 4116–4128. doi: 10.1128/MCB.02210-07
- Xie, J., Mao, Q., Tai, P. W., He, R., Ai, J., Su, Q., et al. (2017). Short DNA hairpins compromise recombinant adeno-associated virus genome homogeneity. *Mol. Ther.* 25, 1363–1374. doi: 10.1016/j.ymthe.2017.03.028
- Yu, L. O. (2009). Bioinformatic analysis of inverted repeats of coronaviruses genome. *Biopolym. Cell* 25, 307–314. doi: 10.7124/bc.0007EA
- Ziebuhr, J. (2004). Molecular biology of severe acute respiratory syndrome coronavirus. *Curr. Opin. Microbiol.* 7, 412–419. doi: 10.1016/j.mib.2004.06.007
- Zuker, M. (2003). Mfold web server for nucleic acid folding and hybridization prediction. *Nucl. Acids Res.* 31, 3406–3415. doi: 10.1093/nar/gkg595

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Bartas, Brázda, Bohálová, Cantara, Volná, Stachurová, Malachová, Jagelská, Porubiaková, Červená and Pečinka. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.