# Evolution of G-quadruplex forming motifs in recent SARS-CoV-2 variants

***Demitra Rooyakkers*, Merit Kayastha, Dr. Paramjeet Bagga, Dr. Scott Frees***

Ramapo College of New Jersey, Mahwah, NJ, 07430

## Introduction

### SARS-CoV-2 Virus

**Severe acute respiratory syndrome coronavirus** is a >30 kb enveloped **positive-sense** single-stranded RNA virus. Upon infection, it undergoes intracellular expression and replication of genomic RNA, generating copies that form viral particles. The genome contains **ORF1a and ORF1ab**, encoding 16 non-structural proteins (**nsps**) which form the replication and transcription complex (**RTC**). These nsps play integral roles in various viral properties such as host immunity modulation, proteolysis, RNA synthesis, proofreading, and modification. The remaining genome encodes essential accessory proteins, prominent determinants of viral pathogenicity.

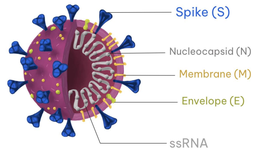



**Figure 1.** SARS-CoV-2 basic structure.

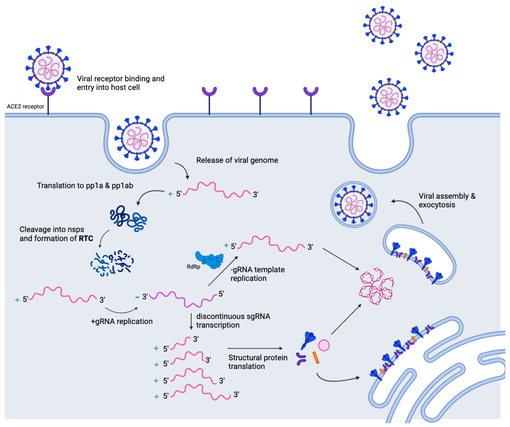



**Figure 2. Intracellular replication cycle of SARS-CoV-2.** Upon entering a host cell, viral genome expression and RNA synthesis are highly regulated. Translation of ORFs generates pp1a and pp1ab polyproteins containing all nsps. Proteolytic nsps, such as nsp3 and nsp5, cleave the polyproteins into separate nsps, forming the RTC. Concurrently, positive-sense genome replication generates a negative-sense RNA template, used to synthesize new positive-sense genomic RNA for virion assembly. Discontinuous transcription of the negative strand produces a nested set of sub-genomic mRNA (sgRNA) for translation into structural and accessory proteins. New virions are assembled in the ER-to-Golgi intermediate compartment and released via exocytosis from infected host cells.

### G-Quadruplexes

A **G-quadruplex** is a four stranded stable DNA structure involved in vital biological processes. It is formed by a repeated folding of a polynucleotide molecule into stacked **G-tetrads**, in which hydrogen bonding connects guanine bases. This base pairing forms a 3-D structure within DNA and RNA molecules that has significant roles in regulation of biological processes and therapeutic targets. **Putative G-quadruplexes** are identified using the following motif:

$$G_xN_{y1}G_xN_{y2}G_xN_{y3}G_x$$

**x** = number of guanine tetrads in the G-quadruplex
$\mathbf{y_1y_2y_3}$ = the length of the loops connecting the guanine tetrads

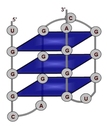


**Figure 3.** G-quadruplex formed by a 'G'-Rich Sequence (QGRS).

## Research Question

Can G-quadruplex motifs potentially influence the replication, transcription, and evolution of SARS-CoV-2?

## Experimental Methods

We have used a computational approach to map and compare QGRS motifs in SARS-CoV-2 variants and identify divergence in data, determining possible implications.

Sequence data

NCBI

GRS

QGRS mapping[7]

EMBL-EBI

Multiple sequence alignments using ClustalW[8]

## 1 Results

### Identities of QGRS Found in Wuhan-Hu-1 Positive Strand

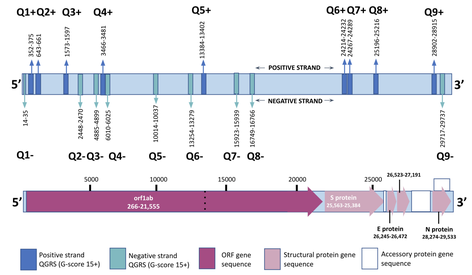
| QGRS | G-Score | Sequence Motif | Reference |
|---|---|---|---|
| *Q1+ | 17 | 352-**GGCTTTGGAGACTCCGTGGAGGAGG**-375 | 1, 2, 3, 6, 10, 11, 12 |
| *Q2+ | 17 | 643-**GGTAATAAAGGAGCTGGTGG**-661 | 1, 3, 6, 10, 11, 12 |
| Q3+ | 19 | 1573-**GGTGTTGTTGGAGAAGGTTCCGAAGG**-1597 | 1, 2, 6, 10, 11 |
| *Q4+ | 17 | 3466-**GGAGGAGGTGTTGCAGG**-3481 | 1, 3, 6, 10, 11 |
| *Q5+ | 19 | 13384-**GGTATGTGGAAAGGTTATGG**-13402 | 1, 3, 6, 10, 12 |
| *Q6+ | 18 | 24214-**GGTTGGACCTTTGGTGCAGG**-24232 | 1, 3, 6, 10, 11 |
| *Q7+ | 19 | 24267-**GGCTTATAGGTTTAATGGTATTGG**-24289 | 1, 2, 3, 6, 10, 11, 12 |
| *Q8+ | 18 | 25196-**GGCCATGGTACATTTGGCTAGG**-25216 | 1, 3, 6, 10, 11, 12 |
| *Q9+ | 19 | 28902-**GGCTGGCAATGGCGG**-28915 | 1, 3, 6, 10, 11, 12 |

### Identities of QGRS Found in Wuhan-Hu-1 Negative Strand

| QGRS | G-Score | Sequence Motif | Reference |
|---|---|---|---|
| Q1- | 15 | 14-**GGAAGGGTCCATTGTTTGGTTGG**-35 | 10, 12 |
| Q2- | 15 | 2448-**GGATGAGTACGGAGATTTTCGGGG**-2470 | 10, 12 |
| Q3- | 18 | 4885-**GGATGGTGTAAGGTGG**-4899 | 10, 12 |
| Q4- | 19 | 6010-**GGTTTGGTTGGTATAGG**-6025 | 10, 12 |
| Q5- | 19 | 10014-**GGTTGGTGGTGTTTGGAGATAGTGG**-10037 | 10, 12 |
| Q6- | 16 | 13254-**GGCAACGGTGTATCTAGTAGGTTTAGG**-13279 | 10, 12 |
| Q7- | 19 | 15923-**GGAAGGAATGGGTCTAGG**-15939 | 10, 12 |
| Q8- | 17 | 16749-**GGATCTGGTGGTGAATTGG** -16766 | 10, 12 |
| Q9- | 15 | 29717-**GGTGGTGTAAAAGTGGCTCCGG**-29737 | 10, 12 |

**We have identified 9 highly conserved Quadruplex forming G-Rich Sequence (QGRS) motifs in each strand of SARS-CoV-2 (Wuhan-Hu-1).** Our results are in agreement with other studies. We used 'QGRS Mapper' previously developed in our lab to map these motifs.[12]
*QGRS experimentally verified to form *in vitro*.

## 2 Wuhan-Hu-1 Genomic Map




**Our data suggests that QGRS motifs may be involved in the regulation of genes involved in the spike and nucleocapsid proteins, replication, and transcription of SARS-CoV-2.** The QGRS motifs on the positive strand were mapped to genes involved in important +gRNA *replication*, *structural* proteins. We also mapped nine QGRS motifs to the negative strand, which is integral in the continuous replication of +gRNA and discontinuous transcription of sub-genomic RNA that encodes viral structural proteins. **G-quadruplexes are known to be involved in regulating the expression of genes, and our data suggests that QGRS motifs may therefore control the rate of replication and transcription processes of viral RNA.**

## 3 Comparisons of 'Q5-' and 'Q9-' QGRS Data in Negative Strand of Wuhan-Hu-1 and Current Variants

```
A Wuhan.      TGGTTGGTGGTGTTTGGAGATAGTGGAGT
  XBB-1.16.   TGGTTGGTGGTGTTTAGAGATAGTGGAGT
  JN.1.       TGGTTGGTGGTGTTTAGAGATAGTGGAGT
  BA.2.86.    TGGTTGGTGGTGTTTAGAGATAGTGGAGT
  EG.5.       TGGTTGGTGGTGTTTAGAGATAGTGGAGT
  XBB-1.5     TGGTTGGTGGTGTTTAGAGATAGTGGAGT
              ********* ************

B Wuhan.      CGGTGGTGTAAAAGTGGCTCCGGTGC
  XBB-1.16.   CGGTGGTGTAAAAGTGG---------
  JN.1.       CGGTGGTGTAAAAGTGG---------
  BA.2.86.    CGGTGGTGTAAAAGTGG---------
  EG.5.       CGGTGGTGTAAAAGTGG---------
  XBB-1.5     CGGTGGTGTAAAAGTGG---------
              *****************
```

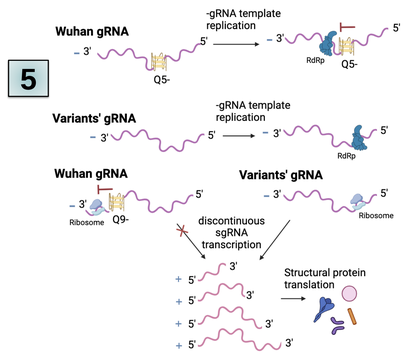
**We next mapped QGRS motifs in current variants of SARS-CoV-2 and found all but 'Q5-' and 'Q9-' QGRS motifs of Wuhan-Hu-1 to be highly conserved among the Wuhan reference genome and the analyzed variants.[8] (A)** A G>A mutation in 'Q5-' QGRS disrupts a G-quadruplex in all the variants tested. This suggests possible evolution over time favoring the loss of 'Q5-'. **(B)** A large deletion in the 'Q9-' QGRS also disrupts a G-quadruplex in all the variants tested. This suggests possible evolution over time favoring the loss of 'Q9-'. G-quadruplex motifs have been known to regulate replication and transcription of viral RNA genomes.[12] **Our data suggests that the G-quadruplex forming sequences present in the Wuhan genome possessed regulatory capabilities which were not favored in the evolution of the viral strains.**

## 4 Validation of QGRS Prediction by Four Different Programs

| QGRS | QGRS Mapper[7] | pqsfinder[5] | G4Catchall[4] | G4Boost[9] |
|---|---|---|---|---|
| Wuhan Q5- | + | + | + | + |
| Variants Q5- | - | + (very weak) | - | - |
| Wuhan Q9- | + | + | + | + |
| Variants Q9- | - | - | - | - |

**In addition to QGRS Mapper, we have used three supplemental G-quadruplex prediction programs to further validate our QGRS comparisons.** The validation of these results strengthens our predictions that the QGRS motifs may form on the negative strand of the Wuhan strain and are not likely to form in the variant strains.[4-5,8-9]

## 5




Putative Mechanism by which G-quadruplexes can impact Replication and Transcription Cycles of SARS-CoV-2

## Conclusions

- The goal of this project was to explore a potential role of G-quadruplexes in the evolution of SARS-CoV-2. We have used a computational approach to map and compare QGRS motifs in variants and identify divergence in data.
- We have identified 9 highly conserved QGRS motifs in each strand of Wuhan-Hu-1.
- Our data suggests that QGRS motifs may be involved in the regulation of genes involved in the spike and nucleocapsid proteins, replication, and transcription of SARS-CoV-2.
- Analysis of recent variants suggests a loss of 'Q5-' and 'Q9-' motifs in current strains, proposing regulatory capabilities which were not favored in the evolution of viral strains.
- We have used three supplemental prediction programs to validate our QGRS findings.
- Previous studies of RNA viruses have confirmed the repressive abilities of G-quadruplexes in the negative strand by reducing the activity of RdRp.[12]
- **Specific ligands can potentially be used to further stabilize the G-quadruplex structures in the viral genome, enhancing their inhibitory properties and serving as plausible targets for antiviral strategies.**

## Literature Cited

1. Bezzi G, Piga EJ, Binolfi A, Armas P. CNBP binds and unfolds in vitro G-quadruplexes formed in the SARS-CoV-2 positive and negative genome strands. *Int J Mol Sci*. 2021; 22(5):2614. https://doi.org/10.3390/ijms22052614
2. Bidula S, Brázda V. Genomic analysis of non-B nucleic acids structures in SARS-CoV-2: potential key roles for these structures in mutability, translation, and replication?. *Genes (Basel)*. 2023;14(1):157.
3. Cui H, Zhang L. G-Quadruplexes are present in human coronaviruses including SARS-CoV-2. *Front Microbiol*. 2020;11(567317). doi:https://doi.org/10.3389/fmicb.2020.567317
4. Doluca O. G4Catchall: A G-quadruplex prediction approach considering atypical features. *J Theor Biol*. 2019;463:92-98.
5. Hon J, Martínek T, Zendulka J, Lexa M. pqsfinder: an exhaustive and imperfection-tolerant search tool for potential quadruplex-forming sequences in R. *Bioinformatics*. 2017;33(21):3373-3379.
6. Ji D, Juhas M, Chi Man Tsang, Chun Kit Kwok, Li Y, Zhang Y. Discovery of G-quadruplex-forming sequences in SARS-CoV-2. *Brief Bioinform*. 2020;22(2):1150-1160.
7. Kikin O, D'Antonio L, Bagga PS. QGRS Mapper: a web-based server for predicting G-quadruplexes in nucleotide sequences. *Nucleic Acids Res*. 2006;34(Web Server issue):W676-W682.
8. Madeira F, Pearce M, Tivey ARN, et al. Search and sequence analysis tools services from EMBL-EBI in 2022. *Nucleic Acids Res*. 2022;50(W1). doi:https://doi.org/10.1093/nar/gkac240
9. Qin G, Zhao C, Liu Y, et al. RNA G-quadruplex formed in SARS-CoV-2 used for COVID-19 treatment in animal models. *Cell Discov*. 2022;8(1). doi:https://doi.org/10.1038/s41421-022-00450-x
10. Rooyakkers D, Kayastha M, Bagga PS, Frees S. Evolution of a G-Quadruplex Forming Sequence in Recent SARS-CoV-2 Variants.; 2024.
11. Zhai LY, Su AM, Liu JF, Zhao JJ, Xi XG, Hou XM. Recent advances in applying G-quadruplex for SARS-CoV-2 targeting and diagnosis: A review. *Int J Biol Macromol*. 2022;221:1476-1490.
12. Zhang R, Xiao K, Gu Y, Liu H, Sun X. Whole genome identification of potential G-quadruplexes and analysis of the G-quadruplex binding domain for SARS-CoV-2. *Front Genet*. 2020;11(587829). doi:https://doi.org/10.3389/fgene.2020.587829