Current Topics in Biochemical Research
Vol. 21, 2020

Original Article

D614G is located in a densely hydrophobic spike glycoprotein motif conserved in the *Sarbecovirus* group

Babu V. Bassa[1,*] and Olen R. Brown[2]

[1]Department of Environmental Toxicology, College of Sciences and Engineering, 108 Fisher Hall, James L. Hunt Street, Southern University and A&M College, Baton Rouge, LA 70813;
[2]Dalton Cardiovascular Research Center, University of Missouri, MO 65211, USA.

ABSTRACT

Severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2) is the pathogen responsible for the COVID-19 pandemic. The D614G substitution appeared in the spike glycoprotein (SGP) of SARS-CoV-2 at an early stage of the COVID-19 outbreak and the mutant carrying the substitution quickly became the most prevalent SARS-CoV-2 variant at several COVID-19 epicenters across the world. There has been a debate on the nature of the mutation, some even suggesting that the mutation is likely to be functionally neutral. One of the ways to understand the nature of this mutation is to study the evolutionary history of the 614th amino acid position of the spike glycoprotein in coronaviruses. In the present bioinformatics study we analyzed a few hundred SARS-CoV (of 2003 outbreak), SARS-CoV-2 and animal SARS-Like strains of the *Sarbecovirus* group to obtain insights into the conservation of aspartic acid at the 614th amino acid position in the SGP of these viruses, using our software tool *Compare*. After analyzing the conservation profiles of several *Sarbecovirus* sequences obtained from GenBank we show here that the 614 amino acid residue is located in an 11-amino acid densely hydrophobic motif, ***vavlyqdvnct*** (11-aa) that is almost perfectly conserved in the *Sarbecovirus* sub-genus and the D614G substitution increased the hydrophobicity of the motif by about 38%. We also identified one SARS-CoV genome in the pre-pandemic GenBank records (Accession: FJ882963) that has the D614G substitution. The fact that the original strain of SARS-CoV-2 with aspartic acid at the 614th position (D-variant) is now almost extinct emphasizes the importance of the D614G mutation for the survival of the virus. We propose that the D614G substitution altered the chemistry of the 11-aa and helped in the spread and continued survival of the virus by some yet to be identified biochemical mechanisms.

KEYWORDS: SARS-CoV-2, D614G, SARS-Like, coronavirus, COVID-19.

INTRODUCTION

The first cases of the present coronavirus pandemic were reported in the Wuhan City of China in December 2020 [1]. The pathogen causing the syndrome was almost immediately identified as a coronavirus strain and it was given the designation SARS-CoV-2 to distinguish it from the coronavirus strain, SARS-CoV, of the 2003 outbreak [2]. The World Health Organization later declared the outbreak a pandemic and named the disease as Coronavirus Disease-19 (COVID -19) [3]. The genomic sequences of human SARS and the animal SARS-like strains of coronaviruses are annotated under the subgenus *Sarbecovirus* in the GenBank records [4]. Coronaviruses of the *Sarbecovirus* group are enveloped viruses that carry positive stranded RNA as the genetic material. The spike glycoprotein (SGP) and the

*Corresponding author: bassa_babu@subr.edu

nucleocapsid protein (NP) are the two structural proteins that are the integral parts of the envelope [5]. The SGP contains many functional domains that are involved in the entry of the virus into mammalian cells. These domains include the receptor binding domain (RBD), the proteolytic cleavage site (PCS), and the heptad-repeating-1 (HR1) domain, and heptad-repeating-2 (HR-2) domains. Whereas the RBD facilitates the initial binding of the virus to the cell, the PCS, HR1, and HR-2 domains are involved in the fusion and entry of the virus into the cells. Genomic mutations in these viruses arise due to RNA replication errors and the sequences carrying the mutation are amplified only if the substitution is beneficial to the virus in terms of its reproductive fitness. Therefore, from this population genetic view-point, genomic mutations once fixed always have a positive effect on the multiplication of the virus [6].

In the present study, in order to understand the probable influence of the D614G substitution on the propagation of the virus, we analyzed several viruses of the *Sarbecovirus* group for SGP mutations using our previously described software tool *Compare* [7]. We learned that D614G is located in an 11-amino acid peptide motif ***vavlyqdvnct*** that is extraordinarily conserved in human SARS and animal SARS-like strains of the *Sarbecovirus* subgenus. These virus strains included bat, civet, and pangolin SARS-like strains, human SARS strains including ExoN 1, wtic-MB, SARS-CoV and SARS-CoV-2, with the exception of the G-variants of SARS-CoV-2 and one G-variant of SARS-CoV (Accession: FJ882963), where the aspartic acid is substituted by glycine in the ***vavlyqdvnct*** motif resulting in ***vavlyqgvnct***. As explained later, we have also identified three GenBank records of previously unreported mutations in the 11-aa of the SARS-CoV-2 strains.

MATERIALS AND METHODS

Software tools and the source of SGP sequences

All the coronavirus spike glycoprotein sequences analyzed in this study were obtained from the publicly maintained database, GenBank. Three software tools namely, Compare [7], Basic Local Alignment Search Tool (BLAST) [8], and Composition-20, were used in this study for the analysis of the sequences. The *Compare* program was written by one of the present authors (Babu V. Bassa). Compare is a multipurpose program which compares biological sequences by identifying domains that are common to the query pair of sequences, each of which is at least three residues long. The program can also be used to query for the presence of shorter peptide motif or parts of it in a larger sequence. *Composition-20* (written and implemented in Visual Studio by one of the present authors) computes amino acid composition and the hydropathy index of any peptide or protein query sequence. The program computes the hydropathy index by the method of Kyte and Doolittle [9].

Identification of the peptide motif (11-aa) that harbors D614G

The 11-aa was first identified by comparing the SGP sequences of SARS-CoV-2 (Accession: **NC_045512**) and SARS-CoV (Accession: **NC_004718**) using *Compare*. Data on the prevalence of the mutations were obtained from GISAID database employing their query options (filters) [10], and the frequencies were calculated as percentages of the total available sequenced genomes for the specified time periods. A sufficient number of samples were analyzed in each case to meet 95% confidence levels statistically.

Analysis for the conservation of 11-AA

Following its identification, the 11-aa was first queried against the SGP sequences of coronavirus strains from the *sarbecovirus* subgenus (n = 91). The 11-aa was then queried in the GenBank database using BLAST.

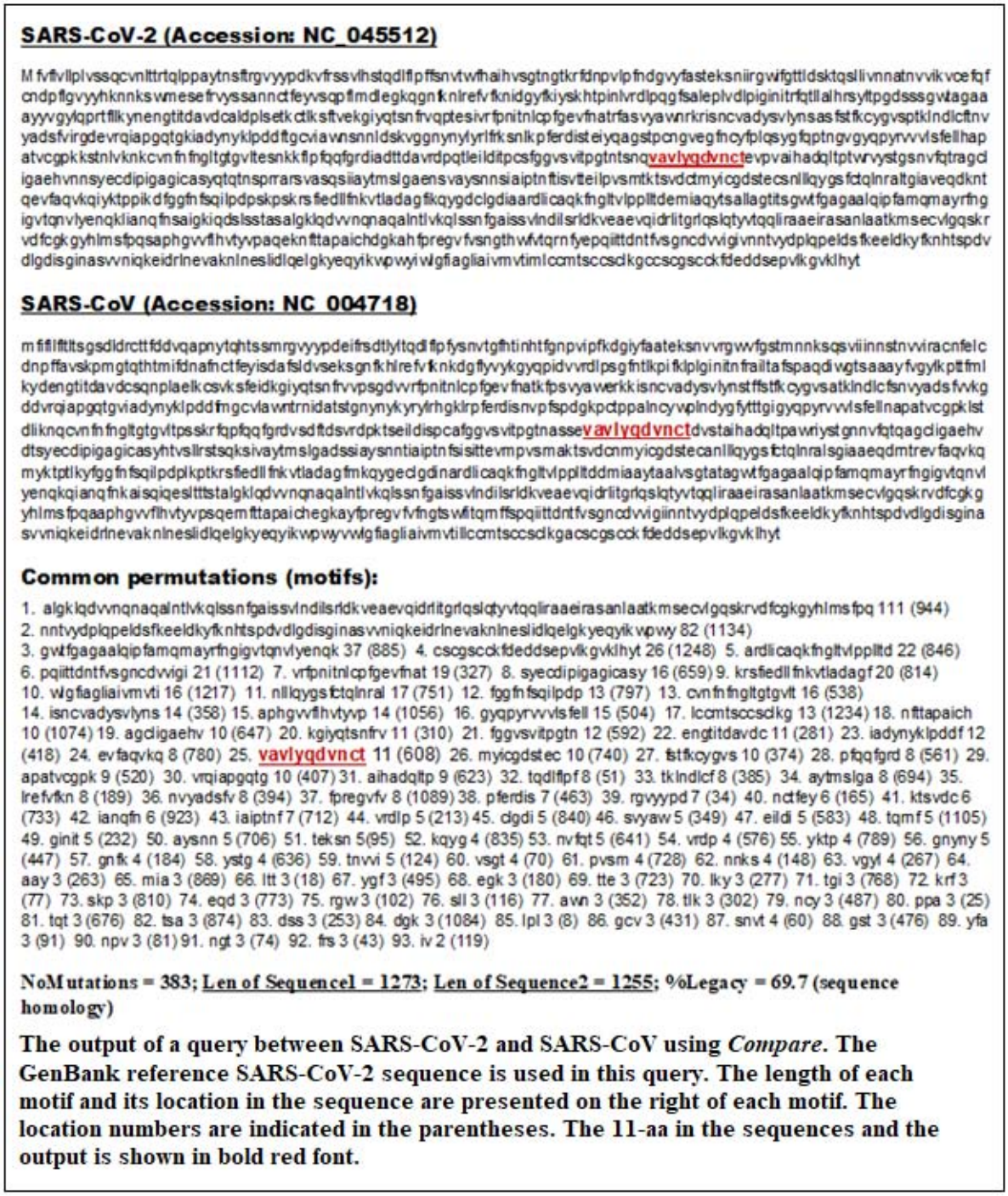
RESULTS AND DISCUSSION

The software tool *Compare* identifies common permutations of three residues and longer occurring between any queried pair of sequences. In our previous report we published a panel of common permutations (conserved motifs) that appeared between SARS-CoV and SARS-Co-V-2 [7]. The peptide motif ***vavlyqgvnct*** was one of the members of that panel. However, the D614G substitution was not very well known at that time.

Upon reinvestigation following the recognition of the D614G mutation [11, 12], we were able to locate the substitution to the ***vavlyqdvnct*** motif (11-aa). There is a significant difference in the sizes of the spike glycoproteins between SARS-CoV (1255 aa) and SARS-CoV-2 (1273 aa) strains of the coronavirus. Investigators have used various criteria in locating the functional domains of the SGP in SARS-CoV-2 [13, 14]. Our identification of the ***vavlyqdvnct*** in SARS-CoV and SARS-CoV-2 strains is based on the assumption that the probability of the 11-aa permutation to have occurred independently in these two spike proteins is infinitely small and that they both originated from a common ancestor. The 11-aa is present at the 608th position in the SGP of SARS-CoV-2 with the aspartic acid (D-variant) or glycine (G-variant) residue being present at the 614th position. The original Wuhan city variant of SARS-CoV-2 has aspartic acid (d) at this position. The 11-aa is present at the 594th position of the SGP in SARS-CoV with "d" being at 600. By scanning the 11-aa against the SGP sequences of *sarbecovirus* strains/isolates we found that the ***vavlyqdvnct*** is almost universally present in all *Sarbecovirus* strains/isolates annotated in the GenBank prior to the COVID-19 pandemic, as an identical permutation with the exceptions of FJ882963 (Human SARS-CoV), where "d" is substituted by "g" and BAE93401 (Human SARS-CoV) where the first amino acid residue valine of 11-aa was substituted with phenylalanine (F). The typical output of the query between SGP of SARS-CoV and SARS-CoV-2 Wuhan strain (D-variant) using *Compare* is presented in Figure 1.

BLAST analysis of ***vavlyqdvnct*** yielded scores of 39.2 and 34.6 respectively for ***vavlyqdvnct*** and ***vavlyqgvnct*** motifs. These scores represent the degree of homology between the query motif and the nearest matching sequences available in the GenBank records. Out of the first 5000 items retrieved in the BLAST query of ***vavlyqdvnct*** motif an estimated 1095 SARS-CoV-2 spike proteins contained the ***vavlyqdvnct*** motif and an estimated 3862 SGP sequences contained the ***vavlyqgvnct*** motif. Please note that the GenBank records contained several records that show "x" at the 614th position of the spike glycoproteins. The GenBank deposits under the *Sarbecovirus* subgenus are now overwhelmed with SARS-CoV-2 genomic sequences and accessing other *Sarbecovirus* genomes through a query with ***vavlyqdvnct*** is now a difficult task. However, our BLAST data accessed in May 2020 confirmed our earlier *Compare* findings that the ***vavlyqdvnct*** motif is present as an identical permutation in all available bat (n = 55), civet (n = 4) and pangolin (n = 6) SARS-like strains of coronavirus. Mutations in this motif appeared only in human SARS-CoV and SARS-CoV-2 strains and the most prominent of those mutations is the D614G. As revealed by *Compare* one of the pre-pandemic GenBank annotations of SARS-CoV contained the D614G substitution (FJ882963) and one other SARS-CoV annotation contained a substitution in 11-aa resulting in the motif ***favlyqdvnct*** (Accession: BAE93401). Both of these strains have human as a host. However, the BLAST query with ***vavlyqdvnct*** revealed two previously unreported substitutions in the 11-aa of SARS-CoV-2 namely, ***vavlyqdinct*** (QN091763) and ***vavlyhdvnct*** (QN091835.1).

The relative positions of 11-aa in SARS-CoV and SARS-CoV-2, respectively, are shown diagrammatically in Figure 2A. The 11-aa is located in close proximity to both the receptor binding domain (RBD) and the protease cleavage site (PCS). The proteolytic cleavage of the SGP into S1 (N-terminal) and S2 (C-terminal) subunits at the cleavage site is believed to enhance the entry of the virus (S-2 subunit together with RNA strand) into the mammalian cells [15]. The 11-aa forms a densely hydrophobic motif. We used the Kyte and Doolittle method through our *Composition 20* software tool for the calculation of total side chain hydrophobicities of the 11-aa, and two 11 amino acid motifs on the N-terminal and the C-terminal sides of 11-aa (Figure 2B). The total amino acid side chain hydrophobicity is expressed as the hydropathy index which is proportional to the hydrophobicity of a given peptide. The D614G substitution increased the hydrophobicity of 11-aa by about 38%. The effect of this change on infectivity and replication of the virus is not clear. However, some reports have claimed that the substitution resulted in an increased shredding of the virus particles from the infected cells [11]. Nevertheless as shown in



SARS-CoV-2 (Accession: NC_045512)

Mfvflvllplvssqcvnlttrtqlppaytnsftrgvyypdkvfrssvlhstqdlflpffsnvtwfhaihvsgtngtkrfdnpvlpfndgvyfasteksniirgwifgttldsktqsllivnnatnvvikvcefqf cndpflgvyyhknnkswmesefrvyssannctfeyvsqpflmdlegkqgnfknlrefvfknidgyfkiyskhtpinlvrdlpqgfsaleplvdlpiginitrfqtllalhrsyltpgdsssgwtagaa ayyvgylqprtfllkynengtitdavdcaldplsetkctlksftvekgiyqtsnfrvqptesivrfpnitnlcpfgevfnatrfasvyawnrkrisncvadysvlynsasfstfkcygvsptklndlcftnv yadsfvirgdevrqiapgqtgkiadynyklpddftgcviawnsnnldskvggnynylyrlfrksnlkpferdisteiyqagstpcngvegfncyfplqsygfqptngvgyqpyrvvvlsfellhap atvcgpkkstnlvknkcvnfnfngltgtgvltesnkkflpfqqfgrdiadttdavrdpqtleildipcsfggvsvitpgtntsnqvavlyqdvnctevpvaihadqltptwrvystgsnvfqtragcl igaehvnnsyecdipigagicasyqtqtnsprrarsvasqsiiaytmslgaensvaysnnsiaiptnftisvteilpvsmtktsvdctmyicgdstecsnlllqygsfctqlnralsgiaveqdknt qevfaqvkqiyktppikdfggfnfsqilpdpskpskrsfiedllfnkvtladagfikqygdclgdiaardlicaqkfngltvlpplltdemiaqytsallagtitsgwtfgagaalqipfamqmayrfng igvtqnvlyenqklianqfnsaigkiqdslsstasalgklqdvvnqnaqalntlvkqlssnfgaissvlndilsrldkveaevqidrlitgrlqslqtyvtqqliraaeirasanlaatkmsecvlgqskr vdfcgkgyhlmsfpqsaphgvvflhvtyvpaqeknfttapaichdgkahfpregvfvsngthwfvtqrnfyepqiittdntfvsgncdvvigivnntvydplqpeldsfkeeldkyfknhtspdv dlgdisginasvvniqkeidrlnevaknlneslidlqelgkyeqyikwpwyiwlgfiagliaivmvtimlccmtsccsclkgccscgscckfdeddsepvlkgvklhyt

SARS-CoV (Accession: NC_004718)

mfiflfttlsgsdldrcttfddvqapnytqhtssmrgvyypdeifrsdtlyltqdlflpfysnvtgfhtinhtfgnpvipfkdgiyfaateksnvvrgwvfgstmnnksqsviiinnstnvviracnfelc dnpffavskpmgtqthtmifdnafnctfeyisdafsldvseksgnfkhlrefvfknkdgflyvykgyqpidvvrdlpsgfntlkpifklplginitnfrailtafspaqdlvgtsaaayfvgylkptlfml kydengtitdavdcsqnplaelkcsvksfeidkgiyqtsnfrvvpsgdvvrfpnitnlcpfgevfnatkfpsvyawerkkisncvadysvlynstffstfkcygvsatklndlcfsnvyadsfvvkg ddvrqiapgqtgviadynyklpddfmgcvlawntrnidatstgnynykyrylrhgklrpferdisnvpfspdgkpctppalncywplndygfytttgigyqpyrvvvlsfellnapatvcgpklst dliknqcvnfnfngltgtgvltpsskrfqpfqqfgrdvsdftdsvrdpktseildispcafggvsvitpgtnassevavlyqdvnctdvstaihadqltpawriystgnnvfqtqagcligaehv dtsyecdipigagicasyhtvsllrstsqksivaytmslgadssiaysnntiaiptnfsisittevmpvsmaktsvdcnmyicgdstecanlllqygsfctqlnralsgiaaeqdmtrevfaqvkq myktptlkyfggfnfsqilpdplkptkrsfiedllfnkvtladagfmkqygeclgdinardlicaqkfngltvlppltddmiaaytaalvsgtatagwtfgagaalqipfamqmayrfngigvtqnvl yenqkqianqfnkaisqiqesltttstalgklqdvvnqnaqalntlvkqlssnfgaissvlndilsrldkveaevqidrlitgrlqslqtyvtqqliraaeirasanlaatkmsecvlgqskrvdfcgkg yhlmsfpqaaphgvvflhvtyvpsqemfttapaichegkayfpregvfvfngtswfitqmffspqiittdntfvsgncdvvigiinntvydplqpeldsfkeeldkyfknhtspdvdlgdisgina svvniqkeidrlnevaknlneslidlqelgkyeqyikwpwyvwlgfiagliaivmvtillccmtsccsclkgacscgscckfdeddsepvlkgvklhyt

Common permutations (motifs):

1. algklqdvvnqnaqalntlvkqlssnfgaissvlndilsrldkveaevqidrlitgrlqslqtyvtqqliraaeirasanlaatkmsecvlgqskrvdfcgkgyhlmsfpq 111 (944)
2. nntvydplqpeldsfkeeldkyfknhtspdvdlgdisginasvvniqkeidrlnevaknlneslidlqelgkyeqyikwpwy 82 (1134)
3. gwtfgagaalqipfamqmayrfngigvtqnvlyenqk 37 (885) 4. cscgsccksfdeddsepvlkgvklhyt 26 (1248) 5. ardlicaqkfngltvlpplttd 22 (846)
6. pqiittdntfvsgncdvvigi 21 (1112) 7. vrfpnitnlcpfgevfnat 19 (327) 8. syecdipigagicasy 16 (659) 9. krsfiedllfnkvtladagf 20 (814)
10. wlgfiagliaivmvti 16 (1217) 11. nlllqygsfctqlnral 17 (751) 12. fggfnfsqilpdp 13 (797) 13. cvnfnfngltgtgvlt 16 (538)
14. isncvadysvlyns 14 (358) 15. aphgvvflhvtyvp 14 (1056) 16. gyqpyrvvvlsfell 15 (504) 17. lccmtsccsclkg 13 (1234) 18. nfttapaich 10 (1074) 19. agcligaehv 10 (647) 20. kgiyqtsnfrv 11 (310) 21. fggvsvitpgtn 12 (592) 22. engtitdavdc 11 (281) 23. iadynyklpddf 12 (418) 24. evfaqvkq 8 (780) 25. vavlyqdvnct 11 (608) 26. myicgdstec 10 (740) 27. fstfkcygvs 10 (374) 28. pfqqfgrd 8 (561) 29. apatvcgpk 9 (520) 30. vrqiapgqtg 10 (407) 31. aihadqltp 9 (623) 32. tqdlflpf 8 (51) 33. tklndlcf 8 (385) 34. aytmslga 8 (694) 35. lrefvfkn 8 (189) 36. nvyadsfv 8 (394) 37. fpregvfv 8 (1089) 38. pferdis 7 (463) 39. rgvyypd 7 (34) 40. nctfey 6 (165) 41. ktsvdc 6 (733) 42. ianqfn 6 (923) 43. iaiptnf 7 (712) 44. vrdlp 5 (213) 45. clgdi 5 (840) 46. svyaw 5 (349) 47. eildi 5 (583) 48. tqmf 5 (1105) 49. ginit 5 (232) 50. aysnn 5 (706) 51. teksn 5(95) 52. kqyg 4 (835) 53. nvfqt 5 (641) 54. vrdp 4 (576) 55. yktp 4 (789) 56. gnyny 5 (447) 57. gnfk 4 (184) 58. ystg 4 (636) 59. tnvvi 5 (124) 60. vsgt 4 (70) 61. pvsm 4 (728) 62. nnks 4 (148) 63. vgyl 4 (267) 64. aay 3 (263) 65. mia 3 (869) 66. ltt 3 (18) 67. ygf 3 (495) 68. egk 3 (180) 69. tte 3 (723) 70. lky 3 (277) 71. tgi 3 (768) 72. krf 3 (77) 73. skp 3 (810) 74. eqd 3 (773) 75. rgw 3 (102) 76. sll 3 (116) 77. avn 3 (352) 78. tlk 3 (302) 79. ncy 3 (487) 80. ppa 3 (25) 81. tqt 3 (676) 82. tsa 3 (874) 83. dss 3 (253) 84. dgk 3 (1084) 85. lpl 3 (8) 86. gcv 3 (431) 87. snvt 4 (60) 88. gst 3 (476) 89. yfa 3 (91) 90. npv 3 (81) 91. ngt 3 (74) 92. frs 3 (43) 93. iv 2 (119)

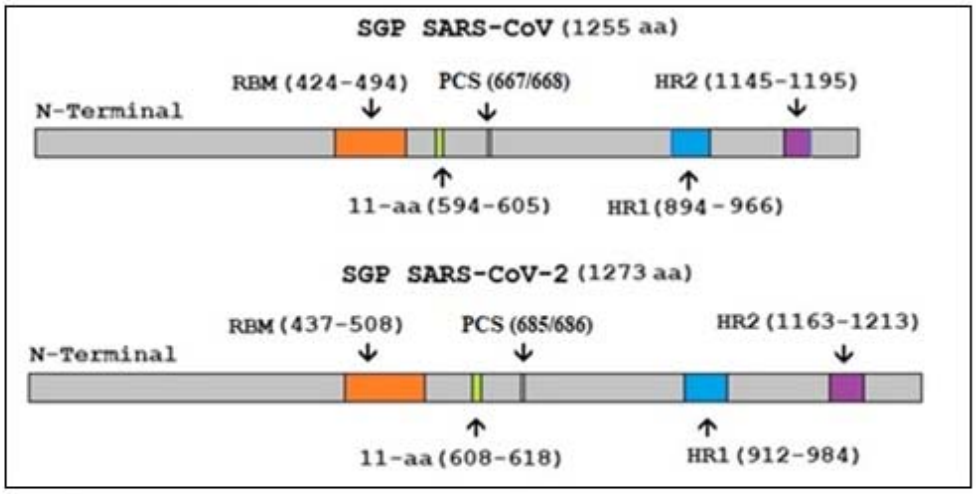
NoMutations = 383; Len of Sequencel = 1273; Len of Sequence2 = 1255; %Legacy = 69.7 (sequence homology)

The output of a query between SARS-CoV-2 and SARS-CoV using *Compare*. The GenBank reference SARS-CoV-2 sequence is used in this query. The length of each motif and its location in the sequence are presented on the right of each motif. The location numbers are indicated in the parentheses. The 11-aa in the sequences and the output is shown in bold red font.

Figure 1. Comparison between the spike glycoproteins of SARS-CoV-2 and SARS-CoV using *Compare*.

Figure 3, the original D-variant that emerged out of Wuhan city, China, spread across the globe, and then disappeared rapidly in the ensuing months. This observation indicates that social distancing measures were able to control the spread of the D-variant but not the G-variant of SARS-CoV-2. Based on the empirical data presented in Figure 3, the D614G mutation is

A. The location of 11-aa in the spike glycoproteins relative to the known functional domains



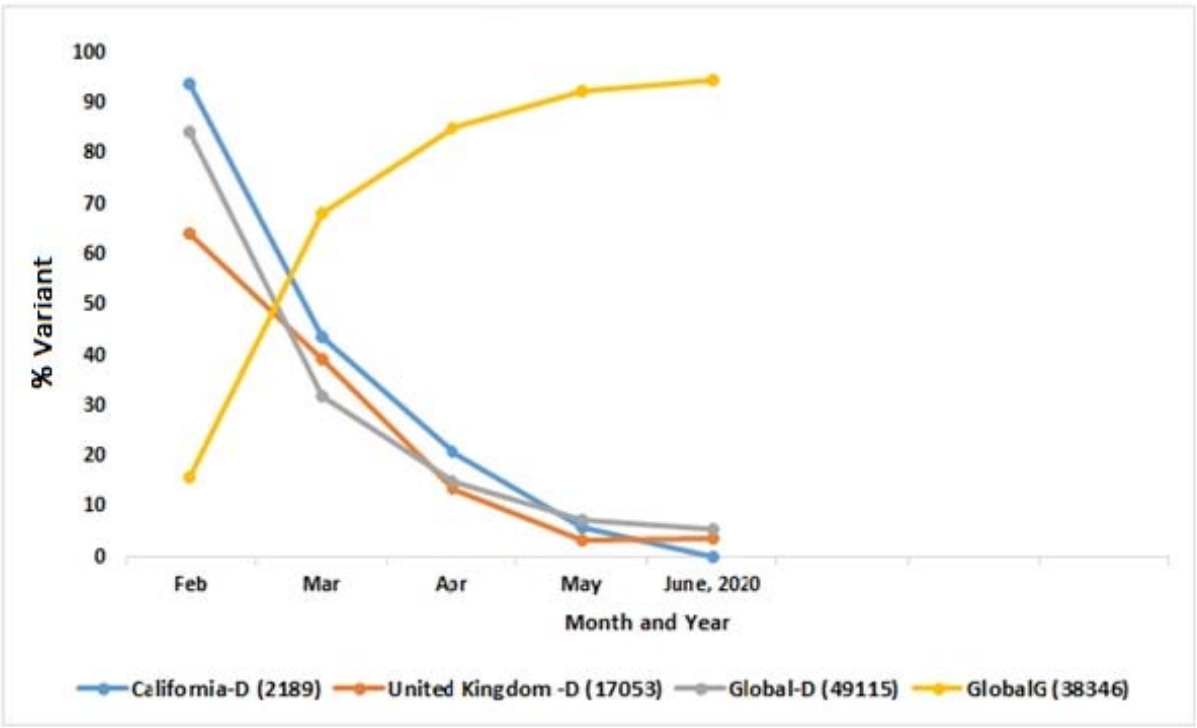
The spike glycoproteins (SGP) of SARS-CoV and SARS-CoV-2 are depicted diagrammatically. The D614G is embedded in the 11-aa. The 11-aa is located in close proximity to both the receptor binding domain (RBM) and the proteolytic cleavage site (PCS). The proteolytic cleavage divides the SGP into S1 (N-terminal) and S2 (C-terminal) subunits. The heptad-repeating-1 (HR1) and heptad-repeating-2 (HR2) domains participate in the fusion of the virus with host cell plasma membrane.

B. The hydropathicity indices of 11-aa and its vicinity in D an G variants

SARS-CoV (D)	-vitpgtnasse	vavlyqdvnct	dvstaihdql-
	(-1.5)	(8.2)	(0.9)
SARS-CoV (G)	-vitpgtnasse	vavlyqgvnct	dvstaihdql-
	(-1.5)	(11.3)	(0.9)
SARS-CoV-2 (D)	-vitpgtntsnq	vavlyqdvnct	evpvaihdql-
	(-6.7)	(8.2)	(5)
SARS-CoV-2 (G)	-vitpgtntsnq	vavlyqgvnct	evpvaihdql-
	(-6.7)	(11.3)	(5)

The 11-aa is highly conserved in the Sarbecovirus group. The hydropathy indices of 11-aa and 11 amino acids preceeding and 11-amino acids following the 11-aa are determined by Kyte and Doolittle method. The D614G mutation increased the hydrophobicity of the 11-aa by approximately 38%. Shown in the parentheses are the hydropathy indices which are proportional to total side-chain hydrophobicities.

Figure 2. A. The location of 11-aa in the spike glycoproteins relative to the known functional domains. B. The hydropathicity indices of 11-aa and its vicinity in D and G variants.



Data from the GISAID database were analyzed using their query language (filters) to retrieve data on the basis of sample collection dates and geographical locations. The D-variant of the SARS-CoV-2 declined rapidly. In the course of the pandemic , the D-variant existed in predominant numbers only in the January, February, and March 2020 samples of China. The G-variant appeared in the Chinese samples as early as January and it was at 26% in the Chinese samples of February, 2020. This indicates the propensity of the G-variant. Near complete extinction of the D-variant is surprising because the D-variant contained the predominant structural changes in the SGP except the D614G substitution. Therefore, our finding that the D614G is located in a highly conserved peptide is significant. It is clear that the G-variant made a major difference in prolongation of the pandemic. By March, 2020 most COVID-19 epicenters across the world carried the G-variant in overwhelming proportions.

Figure 3. The decline of the D-variant and the emergence of the G-variant at various COVID-19 epicenters.

present in the virus that prolonged the COVID-19 pandemic beyond the month of March 2020.

CONCLUSIONS

We have shown that the D614G mutation is embedded in a densely hydrophobic amino acid motif in the spike glycoprotein of the *Sarbecovirus* strains. The motif did not have any substitutions in any of the bat, civet, pangolin, and most human SARS strains until the emergence of SARS-CoV-2. Substitutions, however, appeared in the immediate vicinity of this motif in the *Sarbecovirus* strains. The extraordinary degree of conservation of the 11-aa motif indicates that it played an important role in the replication and survival of the *Sarbecovirus* strains. Thus, it may be presumed that

the D614G substitution could not have established itself unless it had a positive effect on the overall survival of the SARS-CoV-2 strain. We have, however, found one recorded case of the D614G substitution from the GenBank deposits belonging to the 2003 SARS outbreak. This finding is consistent with the human source of this mutation.

The D614G mutation first appeared in the February and March 2020 samples from COVID-19 in China. It is likely that the same mutation emerged also at other geographical locations simultaneously. The frequency of the D-variant has decreased rapidly across the world and the G-variant became established at almost all global COVID-19 epicenters. Based on the frequencies of the D and G variants, much of the morbidity and mortality caused by the pandemic is empirically attributable to the G-variant and less to the D-variant. This observation also means that the D614G mutation most certainly is beneficial to the virus in terms of its reproductive fitness. A change in the chemistry of the 11-aa has more likely contributed to this consequence.

CONFLICT OF INTEREST STATEMENT

There are no conflicts of interest.

REFERENCES

1. Peng, Z., Xing-Lou, Y., Xian-Guang, W., Ben, Hu., Lei, Z., Wei, Z., Hao-Rui, S., Yan, Z., Bei, L., Chao-Lin, H., Hui-Dong, C., Jing, C., Yun, L., Hua, G., Ren-Di, J., Mei-Qin, L., Ying, C., Xu-Rui, S., Xi, W., Xiao-Shuang, Z., Kai, Z., Quan-Jiao, C., Fei, D., Lin-Lin, L., Bing, Y., Fa-Xian, Z., Yan-Yi, W., Geng-Fu, X. and Zheng-Li, S. 2020, Nature, 579(7798), 270-273.
2. Roy, M. A., Christophe, F., Azra, C. G., Christl, A. D., Steven, R., Neil, M. F., Gabriel, M. L., Lam, T. H. and Anthony, J. H. 2004, Phil. Trans. R. Soc. Lond., 359, 1091. doi:10.1098/rstb.2004.1490.
3. www.who.int/news/item/29-06-2020-covid timeline (Accessed on November 23, 2020).
4. National Center for Biotechnology Information (NCBI). Bethesda (MD): National Library of Medicine (US), National Center for Biotechnology Information, 1988, Available from: https://www.ncbi.nlm.nih.gov/
5. Cheng, R. H., Richard, J. K., Norman, H. O., Michael, G. R., Hok-Kin, C., Thomas, J. S. and Timothy, S. B. 1995, Cell, 80, 621.
6. Wright, S. 1942, Bull. Amer. Math. Soc., 48, 223. doi:10.1090/S0002-9904-1942-07641-5. MR 0006700.
7. Babu, B. V. and Brown, O. R. 2020, Front. Biosci. (Landmark Edn.), 25, 1894. doi: 10.2741/4883.
8. https://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE=Proteins
9. Kyte, J. and Doolittle, R. F. 1982, J. Mol. Biol., 157, 105.
10. Elbe, S. and Buckland-Merrett, G. 2017, Global Challenges, 1, 33-46. doi:10.1002/gch2.1018. PMCID: 31565258.
11. Korber, B., Fischer, W. M., Gnanakaran, S., Yoon, H., Theiler, J., Abfalterer, W., Hengartner, N., Giorgi, E. E., Bhattacharya, T., Foley, B., Hastie, K. M., Parker, M. D., Partridge, D. G., Evans, C. M., Freeman, T. M., de Silva, T. I., Sheffield COVID-19 Genomics Group, McDanal, C., Perez, L. G., Tang, H., Moon-Walker, A., Whelan, S. P., LaBranche, C. C., Saphire, E. O. and Montefiori, D. C. 2020, Cell, 182, 1. doi: https://doi.org/10.1016/j.cell.2020.06.043.
12. Bassa, B. and Brown, O. 2020, Preprints, 2020070488; doi:10.20944/preprints202007.0488.v1.
13. Xia, S., Liu, M., Wang, C., Xu, W., Lan, Q., Feng, S., Qi, F., Bao, L., Du, L., Liu, S., Qin, C., Sun, F., Shi, Z., Zhu, Y., Jiang, S. and Lu, L. 2020, Cell. Res., 30, 343.
14. Yingjie, W., Meiyi, L. and Jiali, G. 2020, PNAS, 117, 13967. https://doi.org/10.1073/pnas.2008209117
15. Xia, S., Lan, Q., Su, S., Wang, X., Xu, W., Liu, Z., Zhu, Y., Wang, Q., Lu, L. and Jiang, S. 2020, Sig. Transduct. Target Ther., 5, 92. https://doi.org/10.1038/s41392-020-0184-0