




Infectivity of SARS-CoV-2: there Is Something More than D614G?

Saathvik R. Kannan^{1,2} · Austin N. Spratt^{1,3} · Thomas P. Quinn⁴ · Xiao Heng⁴ · Christian L. Lorson^{1,5} · Anders Sönnerborg^{6,7} · Siddappa N. Byrareddy⁸  · Kamal Singh^{1,5,7,9,10}

Received: 19 August 2020 / Accepted: 26 August 2020 / Published online: 15 September 2020
© Springer Science+Business Media, LLC, part of Springer Nature 2020

Severe Acute Respiratory Syndrome Coronavirus-2 (SARS-CoV-2), causative agent of Coronavirus Disease 19 (COVID-19), has spread throughout the world since its first emergence in Dec. 2019 in Wuhan, China. As anticipated, SARS-CoV-2 evolution is most likely through selection, perhaps conferring enhanced fitness and/or infectivity. Numerous reports detail the molecular identification of D614G mutation in Spike protein (S-protein) (Korber et al. 2020; Koyama et al. 2020; Maitra et al. 2020) and are experimentally characterizing the functional impact upon the viral life cycle (Li et al. 2020; Luthy and Kistler 1989). Regardless of how D614G arose, it is clear that this variant has a distinct phenotype. With the report of D614G, it was hypothesized that viruses containing G614 were more infectious than D614 viruses (Grubaugh et al. 2020; Korber et al. 2020). Experimental evidence quickly confirmed this hypothesis (Korber et al. 2020; Li et al. 2020). However, a report from COVID-19 Genomics UK Consortium (Report #9 - 25th June 2020) showed G614 virus has grown 1.22 times faster than the D614, but the statistical

significance was low, indicating the role of other factors such as mutations in other genes.

To gain insight into the distribution of mutations in SARS-CoV-2 nonstructural proteins (nsps) and structural proteins, we analyzed protein sequences ($n = 7232$) from the United States ($n = 6302$), Europe ($n = 420$), China ($n = 104$), and India ($n = 406$), and determined the mutations with respect to Wuhan-Hu-1 isolate (NCBI Reference Sequence: NC_045512.2). A Circos diagram showing non-synonymous mutations in select viral proteins (nsp8, nsp12, nsp13, nsp14, and S-protein) is shown in Fig. 1a. SARS-CoV-2 has evolved by developing mutations in different viral proteins. While mutations in 5 proteins is shown in Fig. 1a, mutations in other viral proteins were also identified in this analysis. The data also revealed that some mutations were widespread, while other mutations appeared to be restricted geographically. Most strikingly, mutation P323L in nsp12 (an RNA-dependent RNA polymerase or RdRp) and D614G in S-protein co-evolved throughout the world. Other mutations,

Saathvik R. Kannan and Austin N. Spratt contributed equally to this work.

✉ Anders Sönnerborg
anders.sonnerborg@ki.se

✉ Siddappa N. Byrareddy
sid.byrareddy@unmc.edu

✉ Kamal Singh
singhka@missouri.edu

¹ Bond Life Sciences Center, University of Missouri, Columbia, MO 65211, USA

² Hickman High School, Columbia, MO 65203, USA

³ Department of Mathematics, University of Missouri, Columbia, MO 65211, USA

⁴ Department of Biochemistry, University of Missouri, Columbia, MO 65211, USA

⁵ Department of Veterinary Pathobiology, University of Missouri, Columbia, MO 65211, USA

⁶ Department of Molecular Microbiology and Immunology, University of Missouri, Columbia, MO 65211, USA

⁷ Division of Clinical Microbiology, Department of Laboratory Medicine, Karolinska Institute, Huddinge, 14186 Stockholm, Sweden

⁸ Department of Pharmacology and Experimental Neuroscience, University of Nebraska Medical Centre, Omaha, NE 68198, USA

⁹ Shift Pharmaceuticals, Overland Park, KS 66211, USA

¹⁰ Sanctum Therapeutics Corporation, Sunnyvale, CA 94087, USA

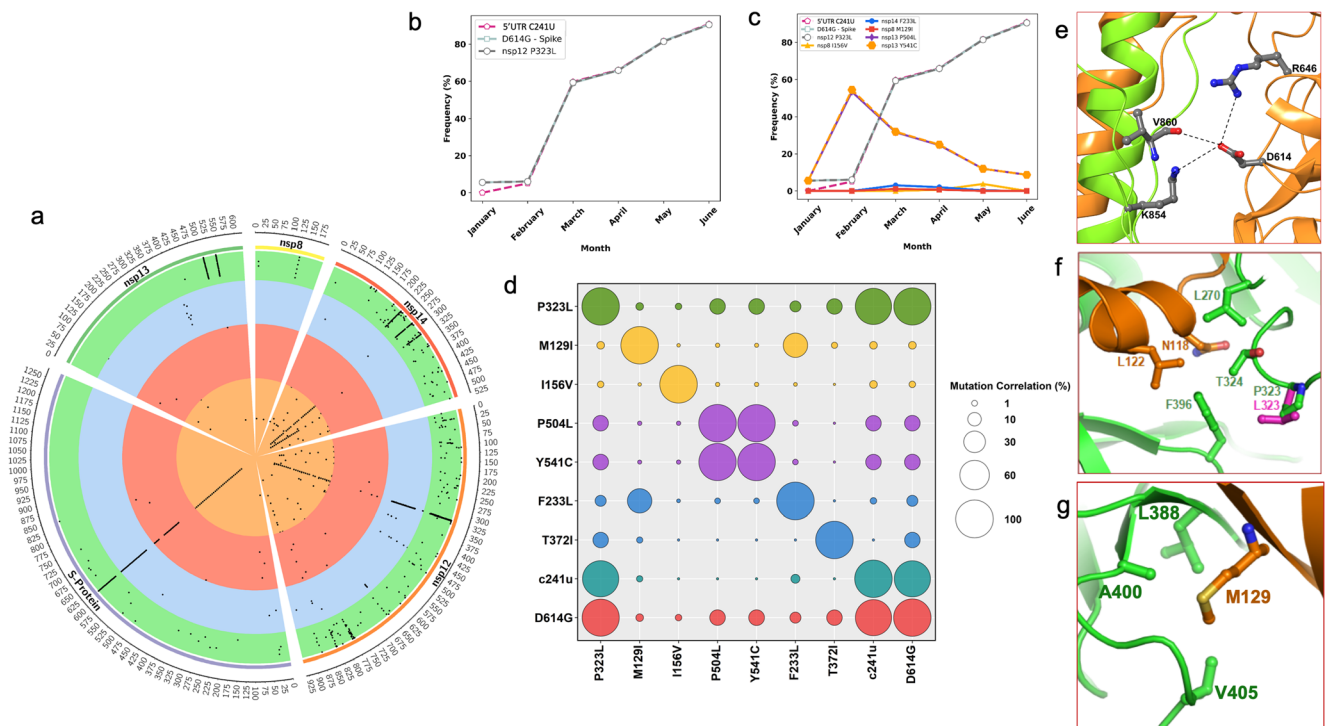


Fig. 1 Details of genetic variations in SARS-CoV-2. **Panel a.** A Circos diagram showing the mutations in nsp8, nsp12, nsp13 and S-protein. The sequences from the USA, Europe, China and India are colored in green, cyan, red and orange, respectively. Each dot represents a mutation in this figure. The Circos diagram was generated by an in-house Circos configuration script (available upon request). **Panel b.** Temporal analysis of mutation frequency of D614G, C241U and P323L in the virus isolated from the USA patients. The frequency was calculated as the number of sequences containing mutation divided by total number of sequences, then multiplied by 100. For example, a T to C mutation frequency = (Number of T mutations / (Number of T mutations + number of regular C)) × 100. **Panel c.** This panel shows temporal change in frequency of I156V and M129I in nsp8 (orange filled triangle and red filled square, respectively), F233L in nsp14 (blue filled circle), P504L and Y541C in nsp13 (violet filled diamonds and orange filled hexagons, respectively) in addition to D614G, P323L and C241U from panel a. Plots in panel b and c were generated by Matplotlib Python script

(available upon request). **Panel d.** Mutual correlation of mutations among different proteins. The normalized mutual correlation was calculated using an in-house python script using scikit-learn (Python) library and plotted with R (codes available upon request). The mutual correlation for T372I was done using a different set of protein sequences ($n = 21$) from India. Final values were multiplied by 100 to express as a percent. **Panel e** shows intra and inter-monomer interactions of D614. These interactions are expected to be lost upon D614G mutation. **Panel f** shows the location of P323 in nsp12 at the interface of nsp8. The hydrophobic residues (sidechain in case of N118 in nsp8) are shown. A mutation P323 to L323 would most certainly enhance the interaction between nsp8 and nsp12 due to increased hydrophobicity of leucine residue. **Panel g** shows the location of M129 within a hydrophobic pocket constituted by L388, A400 and V405 from nsp12. Mutation M129I is expected to enhance the hydrophobic interactions between nsp8 and nsp12. In panel f and g, the green and dark orange colors correspond to nsp12 and nsp8, respectively

such as P504L and Y541C in nsp13 (helicase) were more prevalent in the USA (Fig. 1a).

To further analyze the coevolution of P323L and D614G, we determined the mutation frequency of P323L and D614G in SARS-CoV-2 nucleotide sequences from the United States ($n = 7233$) over a six-month period (January, 2020 - June, 2020). The results of the temporal analysis of the mutation frequency of P323L (nsp12), C241U (5'UTR) and D614G (S-protein) show that P323L was consistently present in the viruses that had D614G mutation and C241U started coevolving with D614G sometime late January 2020 (Fig. 1b). Additionally, there were mutations in other nsps that varied with time. For example, two mutations in nsp13 (P504L and Y541C) coexisted with almost the same frequency over time and these mutations were prevalent up to ~54% by Feb. 2020 before gradually decreasing to ~10% by the end of June 2020

(Fig. 1c). Mutations in nsp7 and nsp8 also emerged over time albeit at low frequency (data not shown).

To examine if a mutual correlation existed among mutations in different proteins and in 5'-untranslated region (5'-UTR) over a six-month period (Jan. 2020 - June 2020), we determined normalized mutual correlation using scikit-learn (a Python program) (Pedregosa et al. 2011). The data shown in Fig. 1d clearly demonstrate that D614G had a near 100% correlation with (i) P323L (nsp12), and (ii) 5'UTR mutation C241U. The other notable correlations were found between M129 (nsp8) and F233L (nsp14) (~15%), P504L/Y541C (nsp13) and D614G /P323L/C241U (~15%), and T372I (nsp14) and D614G/P323L/C241U (~14%) (Fig. 1d).

By the end of June 2020, 91.5% of patients from the U.S. had the D614G mutation (compared to Wuhan-Hu-1 isolate, GenBank ID NC_045512.2). Since SARS-CoV-2 emerged in

Asia, a significant increase in the D614G mutation in the U.S. indicates that this mutation may be associated with enhanced infectivity of the virus outside of China. The structure of G614 S-protein has not been reported yet. However, in the cryoEM structure of the full-length S-protein trimer, D614 is located at the interface of two monomers and forms inter- and intramonomer contacts (Fig. 1e) (Wrapp et al. 2020). D614 forms a salt-bridge with K854 and an H-bond with the backbone of V860 of the neighboring monomer. D614 also forms an intramonomer salt-bridge with R646. A mutation of D614 to G614 should result in the loss of these interactions, which could alter the dynamics of S-protein conformational changes during SARS-CoV-2 infection.

Reported experimental data show that D614G mutation enhances the infectivity of SARS-CoV-2 virus. However, it is highly unlikely that a seemingly high infection rate of SARS-CoV-2 is solely the result of this mutation. A near 100% coexistence of P323L (nsp12) and C241U (5'-UTR) with D614G likely contributes to the viral replication, infectivity, or a combination of attributes that are complex and interplay with the host machinery. P323 is located in the Interface domain (residues 251–398) of nsp12 (Hillen et al. 2020). However, in the cryoEM structure of replication-transcription complex (RTC), consisting of nsp7, nsp8, nsp12 and nsp13, the Interface domain is packed against nsp8 (Chen et al. 2020). The role of nsp7 and nsp8 in the context of the RTC is not known, but they can serve as the processivity factor using the replicating complex of nsp12/nsp7/nsp8/RNA (Hillen et al. 2020; Posthuma et al. 2017) similar to that of thioredoxin in the replicating structure of T7 DNA polymerase (Doublet et al. 1998). P323 is located at the interface of nsp12 and nsp8 (Hillen et al. 2020; Kirchdoerfer and Ward 2019) and mutation P323L is expected to position the leucine side chain within interacting distance to F396, which mediates hydrophobic interactions involving L122 an N118 (C β) of nsp8 and T323 (C γ 2) and L270 of nsp12 (Fig. 1f). Enhanced interaction between nsp12 and nsp8 would further improve the processivity of nsp12 and thereby help in the viral replication. C241 is located in the 5'UTR of the viral RNA, 25 nucleotides upstream of the AUG start codon. SHAPE analysis of viral RNA predicted it in the loop of SL5b (Sun et al. 2020). The SHAPE activity of C241 and neighboring residues in the loop are low in *in vivo* probing than *in vitro* probing, suggesting these residues may be bound by protein or engaged in long-range RNA interactions under *in vivo* conditions, and thus are protected from 2'-OH modification in SHAPE analysis. The C241U may change the structure of the SL5 loop or affect protein binding for easy ribosomal scanning, which is favored in translation initiation.

The nsp8 mutation M129I correlated to D614G at ~2.4%. M129 is also located at the nsp8/nsp12 interface (Fig. 1g). It is positioned such that mutation M129I would result in increased hydrophobic interactions with nsp12 residues L388, A400,

and V405. The increased hydrophobic interactions between nsp12 and nsp8 could also provide enhanced processivity of RNA synthesis by nsp12. Two nsp13 mutations P504L and Y541C are correlated to D614G by ~15%. P504 is at the surface of nsp13, whereas Y541 is located at the putative RNA-binding channel. From the available structures, a reliable role of P504L and Y541C mutations cannot be deduced. Similarly, the two positions (P233 and T372) of nsp14 are away from the 3'-5'-exoribonuclease or methyl-transferase sites. Currently available structure of SARS-CoV nsp14 in complex (Ma et al. 2015) with nsp10 does not provide a conclusive role of mutations at these sites.

In summary, from currently available structural, genetic and biochemical data, the higher infectivity of SARS-CoV-2 D614G mutation is not fully understood. However, from the coexistence of several mutations with relatively high frequency suggest that the infectivity of G614 virus mutation is not solely dependent on the entry. Remarkably, COVID-19 presents with a wide variety of symptoms, including neuropathology, delayed onset symptoms, and anosmia. While these coevolving mutations could impact viral fitness, the breadth and complexity of the clinical symptoms may also be associated with new mutations and adaptations. While there has been an unprecedented eruption of COVID-19 publications, clearly we have much to learn going forward.

Acknowledgments KS acknowledges support from the Office of Research, University of Missouri (Bond Life Sciences Center, Early Concept Grant). AS acknowledges funding from the Swedish Research Council (2020-05836). KS and TPQ acknowledge the computation facilities of the Molecular Interactions Core, University of Missouri.

Author Contributions KS and SNB conceptualized the study; KS wrote the first draft. SRK, ATA, ANS and KS conducted genetic analyses; XH wrote the description on mutation C241U in 5'UTR. SNB, TPQ and CLL, and AS edited the manuscript and contributed to understanding the pathogenicity of CoVs. All authors approved the final manuscript.

References

- Chen J et al. (2020) Structural basis for helicase-polymerase coupling in the SARS-CoV-2 replication-transcription complex cell doi:<https://doi.org/10.1016/j.cell.2020.07.033>
- Doublet S, Tabor S, Long AM, Richardson CC, Ellenberger T (1998) Crystal structure of a bacteriophage T7 DNA replication complex at 2.2 Å resolution nature 391:251-258 doi:<https://doi.org/10.1038/34593>
- Grubaugh ND, Hanage WP, Rasmussen AL (2020) Making sense of mutation: what D614G means for the COVID-19 pandemic remains unclear cell doi:<https://doi.org/10.1016/j.cell.2020.06.040>
- Hillen HS, Kokic G, Farnung L, Dienemann C, Tegunov D, Cramer P (2020) Structure of replicating SARS-CoV-2 polymerase nature:DOI:<https://doi.org/10.1038/s41586-41020-42368-41588>

- Kirchdoerfer RN, Ward AB (2019) Structure of the SARS-CoV nsp12 polymerase bound to nsp7 and nsp8 co-factors *Nat Commun* 10:2342 doi:<https://doi.org/10.1038/s41467-019-10280-3>
- Korber B et al. (2020) Tracking changes in SARS-CoV-2 spike: evidence that D614G increases infectivity of the COVID-19 virus cell doi:<https://doi.org/10.1016/j.cell.2020.06.043>
- Koyama T, Weeraratne D, Snowdon JL, Parida L (2020) Emergence of drift variants that may affect COVID-19 vaccine development and antibody treatment pathogens 9 doi:<https://doi.org/10.3390/pathogens9050324>
- Li Q et al. (2020) The impact of mutations in SARS-CoV-2 spike on viral infectivity and antigenicity cell doi:<https://doi.org/10.1016/j.cell.2020.07.012>
- Luthy RM, Kistler HJ (1989) [visceral leishmaniasis (kala-azar) as an opportunistic infection in immunosuppressed patients] *Schweiz med Wochenschr* 119:1309-1313
- Ma Y et al. (2015) Structural basis and functional analysis of the SARS coronavirus nsp14-nsp10 complex *Proc Natl Acad Sci U S A* 112:9436-9441 doi:<https://doi.org/10.1073/pnas.1508686112>
- Maitra A et al. (2020) Mutations in SARS-CoV-2 viral RNA identified in eastern India: possible implications for the ongoing outbreak in India and impact on viral structure and host susceptibility *J Biosci* 45
- Pedregosa F et al (2011) Scikit-learn: machine learning in Python. *J Mach Learn Res* 12:2825–2830
- Posthuma CC, Te Velthuis AJW, Snijder EJ (2017) Nidovirus RNA polymerases: Complex enzymes handling exceptional RNA genomes *Virus Res* 234:58–73 doi:<https://doi.org/10.1016/j.virusres.2017.01.023>
- Report #9 - 25th June 2020 (2020) COVID-19 Genomics UK (COG-UK) Consortium <https://www.cogconsortium.uk/wp-content/uploads/2020/07/25th-June-2020-Report-COVID-19-Genomics-UK-COG-UK-Consortium.pdf> (Accessed August 15, 2020)
- Sun L et al. (2020) In vivo structural characterization of the whole SARS-CoV-2 RNA genome identifies host cell target proteins vulnerable to re-purposed drugs *BioRxiv*, DOI:<https://doi.org/10.1101/2020.07.07.192732>
- Wrapp D et al. (2020) Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation *science* (New York, NY) 367:1260-1263 doi:<https://doi.org/10.1126/science.abb2507>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.