

SHORT COMMUNICATION

Coronapp: A web application to annotate and monitor SARS-CoV-2 mutations

Daniele Mercatelli¹ | Luca Triboli¹ | Eleonora Fornasari¹ | Forest Ray² | Federico M. Giorgi¹ 

¹Department of Pharmacy and Biotechnology, University of Bologna, Bologna, Italy

²Department of Systems Biology, Columbia University Medical Center, New York City, New York, USA

Correspondence

Federico M. Giorgi, Department of Pharmacy and Biotechnology, University of Bologna, Bologna 40126, Italy.
Email: federico.giorgi@unibo.it

Funding information

Ministero dell'Istruzione, dell'Università e della Ricerca, Grant/Award Number: Montalchini 2016; CINECA, Grant/Award Number: HP10CC5F89

Abstract

The avalanche of genomic data generated from the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) virus requires the development of tools to detect and monitor its mutations across the world. Here, we present a webtool, *coronapp*, dedicated to easily processing user-provided SARS-CoV-2 genomic sequences and visualizing the current worldwide status of SARS-CoV-2 mutations. The webtool allows users to highlight mutations and categorize them by frequency, country, genomic location and effect on protein sequences, and to monitor their presence in the population over time. The tool is available at <http://giorgilab.unibo.it/coronannotator/> for the annotation of user-provided sequences. The full code is freely shared at <https://github.com/federicogiorgi/giorgilab/tree/master/coronannotator>.

KEYWORDS

COVID-19, genetics, mutations, SARS-CoV-2, web application

1 | INTRODUCTION

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is a novel pathogenic enveloped RNA β -coronavirus causing a severe illness in human hosts known as coronavirus disease-2019 (COVID-19). The predominant COVID-19 illness is viral pneumonia, often requiring hospitalization and, in some cases, intensive care.¹ With almost 39 million laboratory-confirmed positive cases worldwide as of October 16, 2020, and an estimated case fatality rate across 204 countries of 5.2%, COVID-19 has become a global health challenge in only a few months.² SARS-CoV-2 infection depends on the recognition of host angiotensin-converting enzyme 2 (ACE2), exposed on the cell surface in human lung tissues.^{3,4} SARS-CoV-2 spike glycoprotein binds ACE2, mediating membrane fusion and cell entry.⁵ Upon cell entry, the virus subverts host cell molecular processes, inducing interferon responses and eventually apoptosis.⁶

To date, much effort has been made to develop therapeutic strategies to limit SARS-CoV-2 transmission and replication, but no treatment or vaccine has proven effective against the virus, and repurposing of approved therapeutic agents has been the main

practical approach to manage the emergency so far.⁷ As viruses mutate during replication, the emergence of SARS-CoV-2 substrains and the challenge of a probable antigenic drift require attention, especially for vaccine development.⁸

Although sequence analyses of SARS-CoV-2 have shown that genomic variability is very low,⁹ new SARS-CoV-2 mutation hotspots are emerging due to the high number of infected individuals across countries and to viral replication rates.¹⁰ Three major SARS-CoV-2 clades known as clade G, V, and S have emerged, showing a different geographical prevalence.¹⁰ The Global Initiative on Sharing All Influenza Data (GISAID) has defined these clades according to the presence of a handful of specific and recurring mutations, highlighted in Table 1. An alternative nomenclature, proposed by the pangolin project,¹¹ is also being used by the community as an alternative to the GISAID one (Table 1).

The most frequent mutation detected so far defines the GISAID G clade and causes an aminoacidic change, aspartate (D) or glycine (G), at position 614 (D614G) of the viral spike protein.¹²

Continual genomic surveillance should be considered to monitor the possible appearance of viral subtypes characterized by altered

TABLE 1 Current nomenclature of SARS-CoV-2 GISAID clades and pangolin lineages by specific mutations (named after genomic coordinate according to reference sequence NC_045512.2)

GISAID clade	Pangolin lineage	Nucleotide mutation	Protein effect
L		Reference genome (NC_045512.2)	
G	B.1	C241T C3037T C14408T A23403G	5'-UTR NSP3:F106F NSP12b:P314L S:D614G
GH	B.1.*	C241T C3037T C14408T A23403G G25563T	5'-UTR NSP3:F106F NSP12b:P314L S:D614G ORF3a:Q57H
GR	B.1.1	C241T C3037T C14408T A23403G GGG28881AAC	5'-UTR NSP3:F106F NSP12b:P314L S:D614G N:RG203KR
S	A	C8782T T28144C	NSP4:S76S ORF8:L84S
V	B.2	G11083T G26144T	NSP6:L37F ORF3a:G251V
O		Others	

Abbreviations: GISAID, Global Initiative on Sharing All Influenza Data; SARS-CoV-2, severe acute respiratory syndrome coronavirus 2; UTR, untranslated region.

tropism or causing more aggressive symptoms. Constant and widespread monitoring of mutations is also a powerful means of informing drug development and global or local pandemic management. GISAID has collected to date (October 16, 2020) over 135,000 publicly accessible SARS-CoV-2 sequences. The GISAID effort has made it possible to compare genomes on a geographical and temporal scale and an increasing number of laboratories have started to sequence COVID-19 patient samples worldwide.^{13,14} Several online tools have been developed to monitor the evolution of the virus from a phylogenetic perspective, such as Nextstrain,¹⁵ or to visualize epidemiological data, such as the number of cases and deaths.¹⁶ However, no online tool currently exists to annotate user-provided SARS-CoV-2 genomic sequences,¹⁷ which may derive from specific GISAID subsets or from sequencing efforts of individual laboratories; neither does any tool specifically monitor the prevalence of specific SARS-CoV-2 mutations associated with particular geographic regions or protein locations nor their frequency in the population over time.

To overcome these limitations, we have developed *coronapp*, a web application with two purposes: real-time tracking of SARS-CoV-2 mutational status and annotation of user-provided viral genomic sequences. Our tool enables users to easily perform

genomic comparisons and provides an instrument to monitor SARS-CoV-2 genomic variance, both worldwide and by uploading custom and locally produced genomic sequences. The webtool is available at <http://giorgilab.unibo.it/coronannotator/> and the full source code is shared on GitHub <https://github.com/federicogiorgi/giorgilab/tree/master/coronannotator>.

2 | MATERIALS AND METHODS

The webtool *coronapp* has been developed using the programming language R and is based on a Shiny server (current version 1.4.0.2) running on R version 3.6.1. The app is based on two distinct files, *server.R* and *ui.R*, managing the server functionalities and the browser visualization processes, respectively. The results visualization utilizes both basic R functions and Shiny functionalities; for tooltip functionality, *coronapp* uses the R package *googleVis* v0.6.4, which provides an interface between R and the Google visualization API.¹⁸

The core of the annotation of the user-provided sequences rests in the NUCMER (Nucleotide Mummer) alignment tool, version 3.1.¹⁹ Nucmer output is processed by UNIX and R scripts provided in GitHub within the *server.R* file.

3 | RESULTS

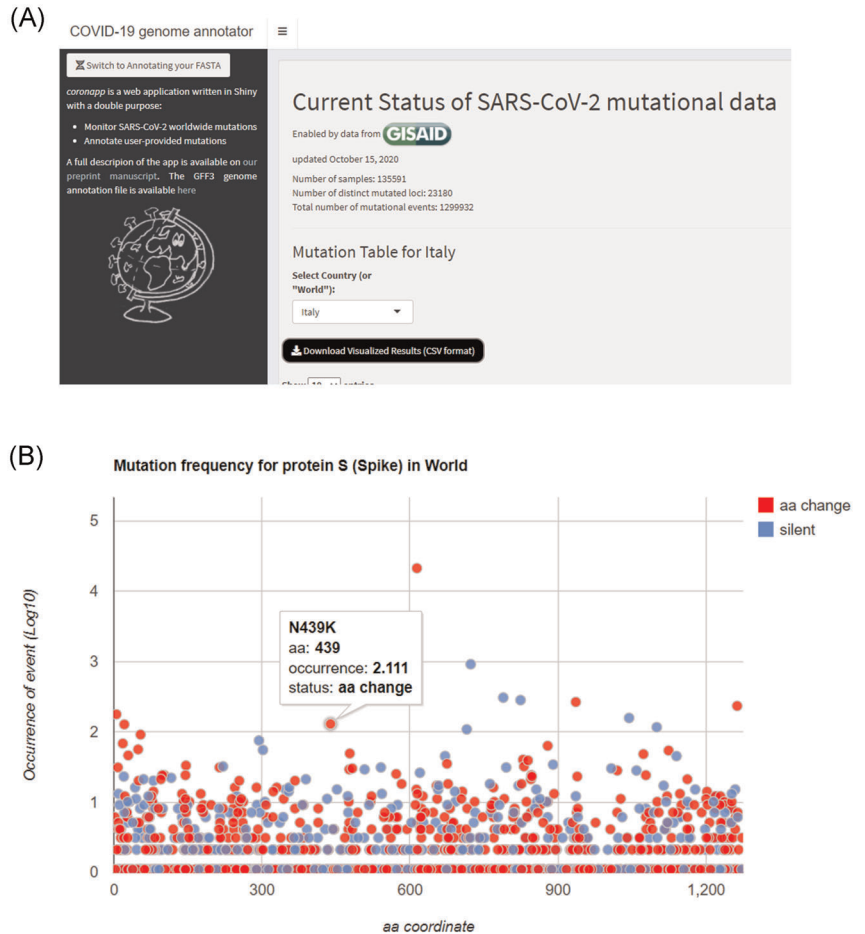
The webtool *coronapp* is available at the website <http://giorgilab.unibo.it/coronannotator/> and it allows the user to annotate SARS-CoV-2 sequences to identify mutations. The app also allows users to annotate user-provided sequences (Figure 1A). There are multiple functionalities of *coronapp*, described in the following paragraphs.

3.1 | Current status of SARS-CoV-2 mutational data

A worldwide analysis is shown, generated using data from GISAID. Specifically, we processed all SARS-CoV-2 complete (>29,000 sequenced nucleotides) genomic sequences, excluding low-quality sequences (>5% undefined nucleotide "N") and viruses extracted from nonhuman hosts.

The underlying database is updated weekly, and we provide the date of the last version as a reference for studies based on the data provided. We indicate the number of samples processed and the total number of mutational events detected (Figure 1A). We also show the number of distinct mutated loci. Currently, this number is slightly below 20,000, meaning that two-thirds of the original Wuhan SARS-CoV-2 genome has been affected by mutations and/or sequencing errors (the full length of the reference genome is 29,903 nucleotides, based on sequence id NC_045512.2). The sequence NC_045512.2 has been, since January 2020, the official NCBI/RefSeq SARS-CoV-2 genome reference. It was originally submitted by

FIGURE 1 Overview of coronapp. (A) Screenshot of the entry page of coronapp showing the basic tool description, the interface to upload user-provided sequences and the overall summary of the mutations detected worldwide. (B) Common interface showing mutation frequency in SARS-CoV-2 proteins, with the occurrence of each mutation on the Y-axis and its corresponding protein coordinate on the X-axis. Red dots indicate aa-changing mutations, and blue dots indicate silent mutations. Tooltip functionality is also provided to identify and quantify each mutation on mouse-over. aa, amino acid; SARS-CoV-2, severe acute respiratory syndrome coronavirus 2



the Shanghai Public Health Clinical Center & School of Public Health (China) on January 5, 2020, and it represents the original SARS-CoV-2 virus from the first outbreak in Wuhan, and annotated as “Wuhan seafood market pneumonia virus”. *Coronapp*, like the majority of current phylogenetic and genomic tools and scientific studies, uses this sequence as the de facto reference of SARS-CoV-2.

3.2 | Mutation frequency in SARS-CoV-2 proteins

We show the frequency of mutations along the length of every SARS-CoV-2 protein, reporting in the X axis the amino acid position and on the Y axis its frequency, either as the number of observed samples carrying the mutation, the base 10 logarithm of that number, or the percentage over all sequenced samples. In the example in Figure 1B, we show the most frequent mutations affecting the viral spike protein S, distinguishing silent mutations and amino acid-changing mutations (including the introduction of STOP codons). For spike, the mutations appear to be evenly distributed in frequency along the protein length, with the most frequent mutation being the aforementioned D614G. Mouse-over functionality is provided to allow the user to identify the selected mutation (e.g., N439K in Figure 1B).

3.3 | The SARS-CoV-2 mutation table

The user can visualize or download the full table of mutations on which the webtool operates (Figure 2A). This table is frequently updated and allows the user to specify a worldwide or a country-specific data set. The table also provides a Search function to look for specific variants or sample ids, and it can be viewed online or downloaded in full as a Comma-Separated Values (CSV) file.

The table shows every mutation in a specific geographical area, reporting:

- The GISAID sample ID (useful for cross-reference with the GISAID database and other analyses based on it, e.g., Nextstrain).
- The country where the sample was collected.
- The position of the mutation, on the reference genome (refpos) and on the sample (qpos).
- The sequence at the mutation site, on the reference genome (refvar) and on the sample (qvar).
- The length of the sample genome (qlength); the reference genome is 29,903 nucleotides long.

(A) Showing results for World

Download Full table (CSV format)

Show 10 entries Search:

sample	country	refpos	refvar	qvar	qpos	qlength	protein	variant	varclass	annotation	varname
EPI_ISL_415706	Switzerland	4	A	T	4	29903	5'UTR	4	extragenic		5'UTR:4
EPI_ISL_415706	Switzerland	241	C	T	241	29903	5'UTR	241	extragenic		5'UTR:241
EPI_ISL_415706	Switzerland	3037	C	T	3037	29903	NSP3	F106F	SNP_silent	Predicted phosphoesterase, papain-like proteinase	NSP3:F106F
EPI_ISL_415706	Switzerland	14408	C	T	14408	29903	NSP12b	P314L	SNP	RNA-dependent RNA polymerase, post-ribosomal frameshift	NSP12b:P314L
EPI_ISL_415706	Switzerland	15324	C	T	15324	29903	NSP12b	N619N	SNP_silent	RNA-dependent RNA polymerase, post-ribosomal frameshift	NSP12b:N619N
EPI_ISL_415706	Switzerland	23403	A	G	23403	29903	S	D614G	SNP	Spike	S:D614G
EPI_ISL_416497	France	4	A	T	4	29862	5'UTR	4	extragenic		5'UTR:4
EPI_ISL_416497	France	241	C	T	241	29862	5'UTR	241	extragenic		5'UTR:241
EPI_ISL_416497	France	2416	C	T	2416	29862	NSP2	Y537Y	SNP_silent	Non-Structural protein 2	NSP2:Y537Y
EPI_ISL_416497	France	3037	C	T	3037	29862	NSP3	F106F	SNP_silent	Predicted phosphoesterase, papain-like proteinase	NSP3:F106F

Showing 1 to 10 of 192,208 entries Previous 1 2 3 4 5 ... 19221 Next

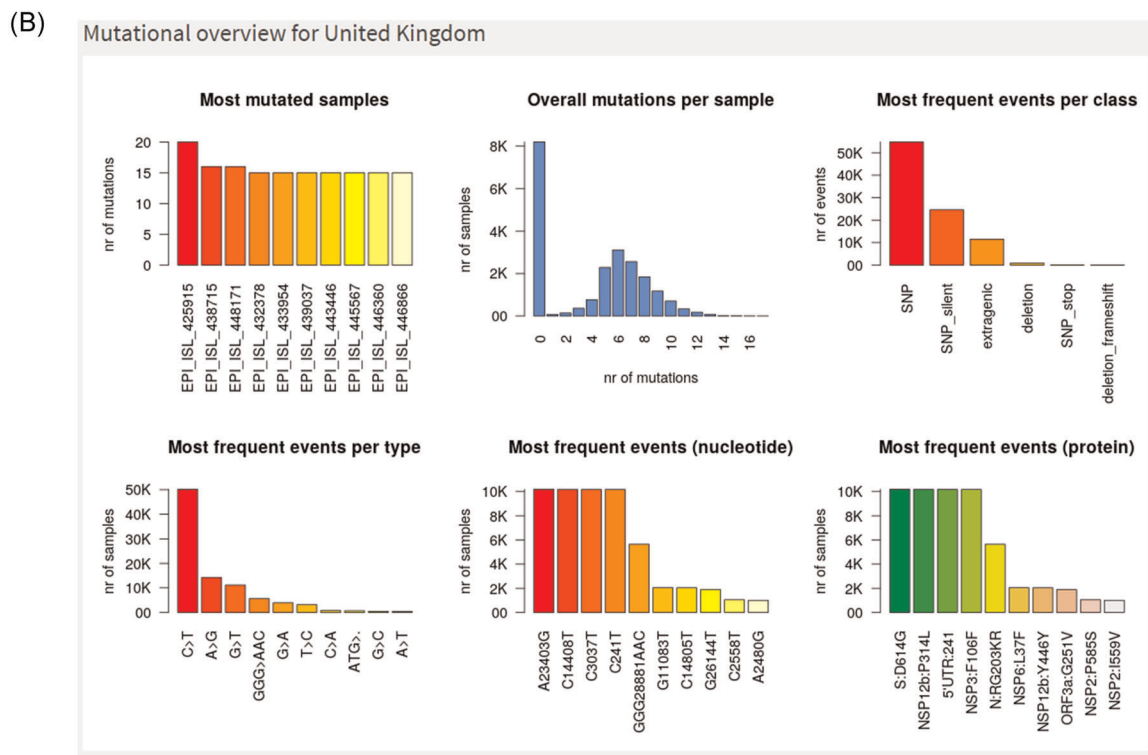


FIGURE 2 Mutation table and overview in coronapp. (A) Result table of coronapp, available both for worldwide-precomputed and user-input analyses. A “download full table” button is provided to allow the user to perform larger-scale analyses autonomously. (B) Bar plots showing the most mutated samples, overall sample mutations and most frequent mutation events, classes, and types. This analysis is also available both for worldwide-precomputed and user-input analyses

- The protein affected by the mutation or, if the mutation is extragenic, the denomination of the untranslated region (UTR), for example, 5'-UTR or 3'-UTR.
- The effect of the mutation on the amino acid sequence of the protein (variant). This uses the canonical mutational standard, indicating the original amino acid(s), the position on the protein,

and the mutated amino acid(s). An asterisk (*) indicates a STOP codon, while the letters indicate amino acids in IUPAC code. For example, a mutation P315L indicates a leucine mutation (L) on the amino acid location 315, normally occupied by a proline (P). Nucleotide mutations can be silent, that is, not yielding any aminoacidic change, for example, the mutation F106F, where the

codon of phenylalanine 106 is affected but without changing the corresponding amino acid. As in the previous column, mutations affecting UTR regions are simply reported as the location of the nucleotide affected.

- The class of the mutation, of which there are currently 10 types:
 - o SNP: a change of one or more nucleotides, determining a change in amino acid sequence.
 - o SNP_stop: a change of one or more nucleotides, yielding the generation of one or more STOP codons.
 - o SNP_silent: a change of one or more nucleotides with no effect in a protein sequence.
 - o Insertion: the insertion of three (or multiples of three) nucleotides, causing the addition of one or more amino acids to the protein sequence.
 - o Insertion_stop: the insertion of three (or multiples of three) nucleotides, causing the generation of a novel stop codon.
 - o Insertion_frameshift: the insertion of nucleotides not as multiples of three, causing a frameshift mutation.
 - o Deletion: the deletion of three (or multiples of three) nucleotides, causing the removal of one or more amino acids to the protein sequence.
 - o Deletion_stop: the removal of three (or multiples of three) nucleotides, causing the generation of a novel STOP codon.
 - o Deletion_frameshift: the deletion of nucleotides not as multiples of three, causing a frameshift mutation.
 - o Extragenic: a mutation affecting intergenic or UTR regions.
- The extended annotation of the protein region affected by the mutation (e.g., “Spike” for “S” or “Predicted phosphoesterase, papain-like proteinase” for NSP3, the nonstructural protein 3).
- The full name of the variant (varname), in the format proteinName:AApositionAA, to allow for a unique denomination of viral proteome variants.

3.4 | Mutational overview

The user is also provided with a general overview of the mutational status of the selected country or the entire world (Figure 2B). Six bar plots provide a summary and highlights of the data set, specifically:

- The most mutated samples, indicating which samples (in GISAID IDs) carry the highest number of mutations.
- The overall mutations per sample, indicating the distributions of mutations per sample. It has been previously reported¹⁰ that the current mode for mutation number compared to the reference NC_045512.2 genome is 7.5.
- The most frequent events per class. Classes are the same as reported in the mutation table and are described in the previous paragraph.
- The most frequent events per type. Individual mutation types are shown as specific nucleotides events, for example, cytosine to thymidine transitions (C>T), guanosine to thymidine transversion (G>T), or even multinucleotide mutations (e.g., GGG>AAC, observed in the nucleocapsid protein). As reported before,

nucleotide transitions seem to be the most abundant SARS-CoV-2 type of mutational event worldwide.¹²

- The most frequent events, either in nucleotide coordinates or in aminoacidic coordinates. Currently, the most frequent events are four mutations affecting SARS-CoV-2 genomes belonging to clade G, which is the most sequenced worldwide and predominant in Europe. These mutations are A23403G (associated with the already mentioned D614G mutation in the spike protein), C3037T, C14408T, and C241T.

3.5 | Analysis of mutations over time

The *coronapp* webtool allows users to monitor the abundance and frequency of any SARS-CoV-2 mutation in any country specified (Figure 3A). Both plots in this section report continuous dates on the X axis, starting on the day of the first collected SARS-CoV-2 genome available on GISAID: December 24, 2019.

The “abundance” plot reports on the Y axis the number of samples carrying a selected mutation in a particular day, in the specified country or worldwide. Since the date reported is the collection date (not the submission date to the GISAID database), there is usually a drop toward the right part of the plot, as there are fewer sequences collected approaching the day of the analysis. The “frequency” plot, on the other hand, normalizes the abundance of mutations by the total number of sequences generated on each day. The plot currently shows a sharp increase in clade G-associated mutations (e.g., S:D614G), as these mutations are most frequent in countries where sequencing is more pervasive (e.g., UK).

3.6 | Annotation of user-provided SARS-CoV-2 genomic sequence

Coronapp provides the user with the possibility of uploading one or more SARS-CoV-2 genomic sequences, which can be complete or partial. The format of the sequences is standard FASTA, and an example input FASTA containing 12 sequences is provided (Figure 3B). The analysis is almost instantaneous and shows an overall breakdown of the most mutated samples and most frequent mutations in the data set. Moreover, a full table of all detected mutations is provided: this can be visualized and searched on the web browser or downloaded as a standard CSV file. Finally, a mutation frequency plot is provided, allowing the user to visualize mutation frequency in selected proteins.

The user can easily return to the worldwide status of the app by refreshing or reopening the page.

4 | DISCUSSION

Our webtool *coronapp* provides a fast, simple tool to annotate user-provided SARS-CoV-2 genomes and visualize all mutations currently present in viral sequences collected worldwide. The results provided

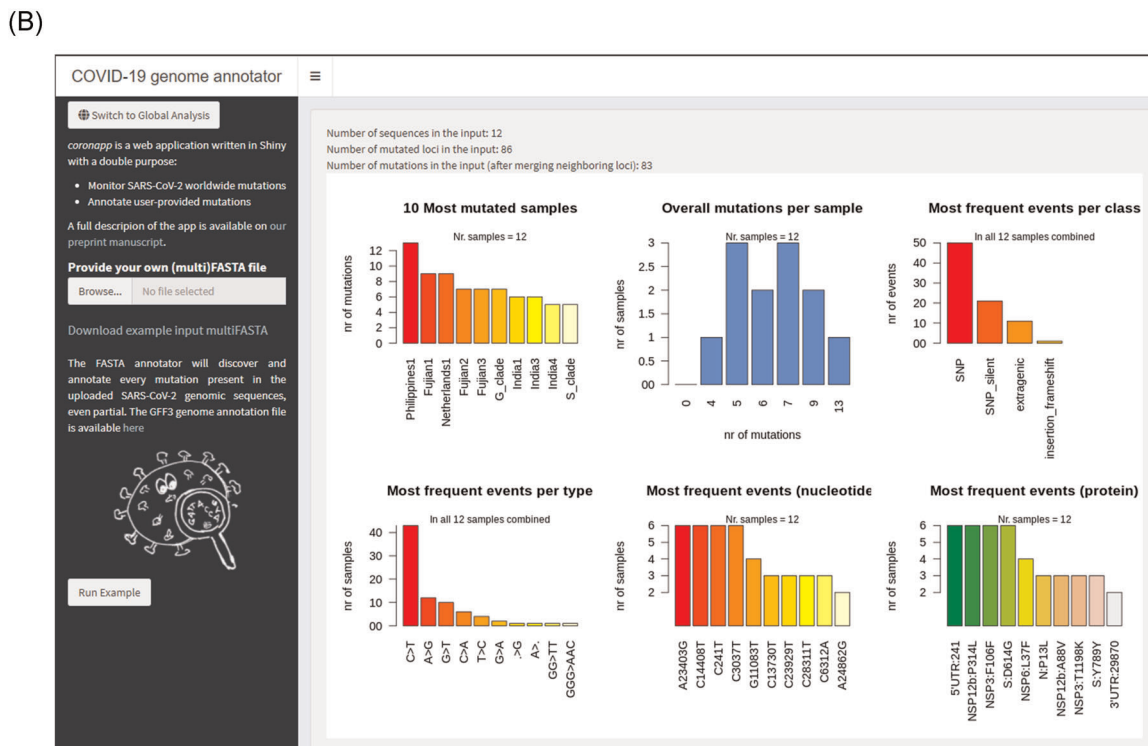
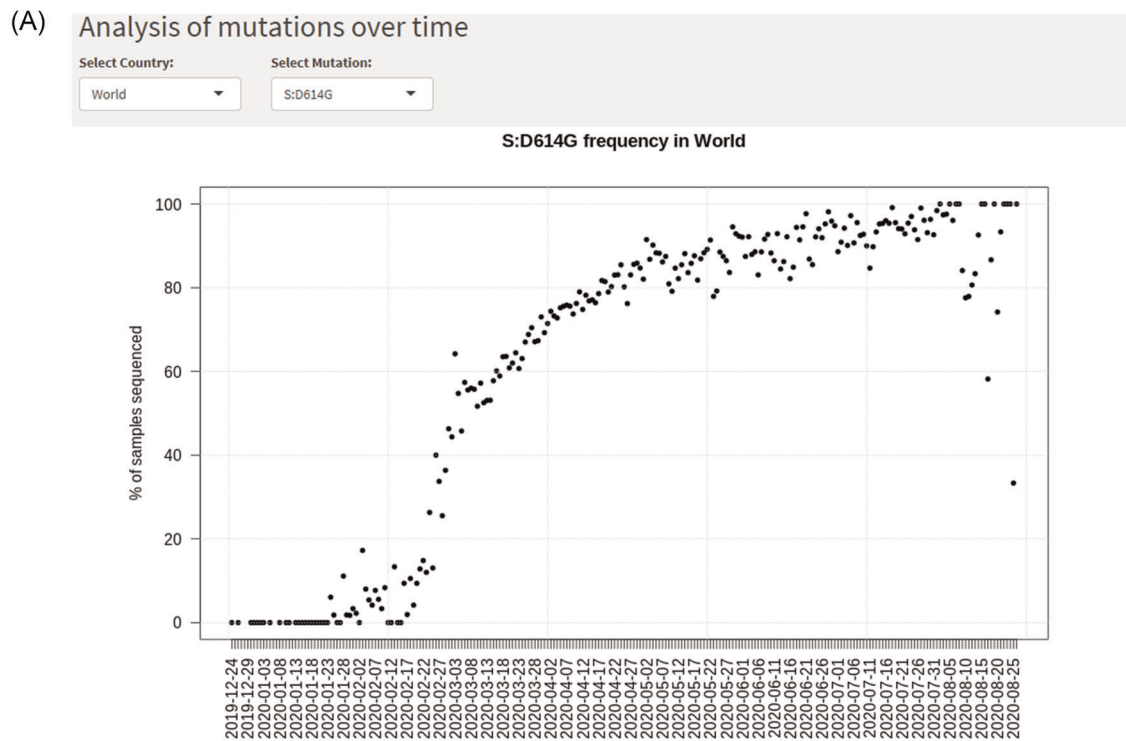


FIGURE 3 Analysis of mutations over time. (A) Chronological analysis made by coronapp, showing the frequency of each user-specified mutation in any user-specified country (or worldwide). The graph shows the same data normalized by the total number of samples, as the percentage of samples sequenced in a specific day and carrying the mutation. (B) Screenshot of the coronannotator companion tool, allowing users to annotate their own SARS-CoV-2 sequences (provided as FASTA files). SARS-CoV-2, severe acute respiratory syndrome coronavirus 2

by this instrument can have several applications. The main purpose of *coronapp* is to help medical laboratories at the front lines of COVID-19 fight with the opportunity to quickly define the mutational status of their sequences, even without dedicated bioinformaticians.

Additionally, it enables scientists to perform mutational covariance analyses and to identify present and future significant functional interactions between viral mutations, as previously attempted for the influenza virus and the human immunodeficiency virus (HIV).²⁰ Another application is the identification of the most frequent mutations in specific protein regions: for example, our tool can quickly identify that the most frequent mutation in the spike protein, D614G, lies outside the known interaction domain with the human protein ACE2, which spans roughly between spike amino acids 330 and 530.²¹ Scientists are currently debating on whether the increased frequency of D614G genomes¹⁰ is the result of SARS-CoV-2 adaptation to the human host²² or rather a mutation with no effective phenotype.²³ A recently published structural model simulating the effect of the D614G mutation on the 3D structure of the spike protein has suggested that this mutation may result in a viral particle, which binds ACE2 receptors less efficiently, due to the masking of the host receptor-binding site on viral spikes.²⁴ The same researchers have reported a possible correlation of the D614G form with increased case fatality rates, hypothesizing that this mutation may lead to a viral form which is better suited to escape immunologic surveillance by eliciting a lower immunologic response.²⁴

The *coronapp* analysis highlighted in Figure 1B shows that a mutation located within the spike/ACE2 interaction domain is the change of asparagine (N) to a lysine (K) in position 439 of the spike sequence; this mutation could affect the protein folding or its affinity with ACE2, as asparagine is less charged than the basic amino acid lysine.

One of *coronapp's* key strengths is to help prioritize scientific efforts on specific aminoacidic variations that could affect the efficacy of antiviral strategies or the development of a vaccine by tracking the most frequent mutations in the population. A further novelty of *coronapp* is that it provides a means to assess the growth or decline of specific mutations over time to identify possible viral adaptation mechanisms.

We provide not only the webtool but also all the underlying code for the annotation and visualization steps on a public GitHub repository, to help other computational scientists in the ongoing battle against COVID-19. Furthermore, the *coronapp* structure and concept could be expanded to other current and future pathogens as well (e.g., seasonal influenza or HIV), to monitor the mutational status across proteins, countries, and time.

ACKNOWLEDGMENTS

We thank the Italian Ministry of University and Research for their support (grant Montalcini 2016) and CINECA (grant HP10CC5F89).

CONFLICT OF INTERESTS

The authors declare that there are no conflict of interests.

AUTHOR CONTRIBUTIONS

Daniele Mercatelli drafted the manuscript and performed the mutational analysis and literature search. Luca Triboli drafted the methodological parts of the manuscript. Eleonora Fornasari worked on the graphical interface of the webtool. Forest Ray wrote the manuscript and performed a literature search. Federico M. Giorgi designed the study, developed the server code and user interface, finalized the manuscript, and provided financial support. All authors tested the webtool and provided original contributions to its development. All authors read and approved the final manuscript.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study were generated by all the contributors of the GISAID consortium, available at <https://www.gisaid.org/>.

ORCID

Federico M. Giorgi  <https://orcid.org/0000-0002-7325-9908>

REFERENCES

- Guan W, Ni Z, Hu Y, et al. Clinical characteristics of coronavirus disease 2019 in China. *N Engl J Med*. 2020;382:1708-1720. <https://doi.org/10.1056/NEJMoa2002032>
- Phua J, Weng L, Ling L, et al. Intensive care management of coronavirus disease 2019 (COVID-19): challenges and recommendations. *Lancet Respir Med*. 2020;8:506-517. [https://doi.org/10.1016/S2213-2600\(20\)30161-2](https://doi.org/10.1016/S2213-2600(20)30161-2)
- Zhang H, Penninger JM, Li Y, Zhong N, Slutsky AS. Angiotensin-converting enzyme 2 (ACE2) as a SARS-CoV-2 receptor: molecular mechanisms and potential therapeutic target. *Intensive Care Med*. 2020;46:586-590. <https://doi.org/10.1007/s00134-020-05985-9>
- Guzzi PH, Mercatelli D, Ceraolo C, Giorgi FM. Master regulator analysis of the SARS-CoV-2/human interactome. *J Clin Med*. 2020;9:982. <https://doi.org/10.3390/jcm9040982>
- Ou X, Liu Y, Lei X, et al. Characterization of spike glycoprotein of SARS-CoV-2 on virus entry and its immune cross-reactivity with SARS-CoV. *Nat Commun*. 2020;11:1620. <https://doi.org/10.1038/s41467-020-15562-9>
- Blanco-Melo D, Nilsson-Payant BE, Liu W-C, et al. Imbalanced host response to SARS-CoV-2 drives development of COVID-19. *Cell*. 2020;181:1036-1045. <https://doi.org/10.1016/j.cell.2020.04.026>
- Tu Y-F, Chien C-S, Yarmishyn AA, et al. A review of SARS-CoV-2 and the ongoing clinical trials. *Int J Mol Sci*. 2020;21(1):2657. <https://doi.org/10.3390/ijms21072657>
- Koyama T, Weeraratne D, Snowdon JL, Parida L. Emergence of drift variants that may affect COVID-19 vaccine development and antibody treatment. *Pathog Basel Switz*. 2020;9(5):324. <https://doi.org/10.3390/pathogens9050324>
- Ceraolo C, Giorgi FM. Genomic variance of the 2019-nCoV coronavirus. *J Med Virol*. 2020;92:522-528. <https://doi.org/10.1002/jmv.25700>
- Mercatelli D, Giorgi FM. Geographic and genomic distribution of SARS-CoV-2 mutations. *Front Microbiol*. 2020;11:1800. <https://doi.org/10.3389/fmicb.2020.01800>
- Rambaut A, Holmes EC, Hill V, et al. A dynamic nomenclature proposal for SARS-CoV-2 to assist genomic epidemiology. *Nat Microbiol*. 2020;5:1403-1407. <https://doi.org/10.1038/s41564-020-0770-5>
- Chiara M, Horner DS, Gissi C, Pesole G. Comparative genomics suggests limited variability and similar evolutionary patterns between major clades of SARS-CoV-2. *BioRxiv*. 2020. <https://doi.org/10.1101/2020.03.30.016790>

13. Gudbjartsson DF, Helgason A, Jonsson H, et al. Spread of SARS-CoV-2 in the Icelandic Population. *N Engl J Med*. 2020;382:2302-2315. <https://doi.org/10.1056/NEJMoa2006100>
14. Fauver JR, Petrone ME, Hodcroft EB, et al. Coast-to-coast spread of SARS-CoV-2 during the early epidemic in the United States. *Cell*. 2020;181:990-996. <https://doi.org/10.1016/j.cell.2020.04.021>
15. Hadfield J, Megill C, Bell SM, et al. Nextstrain: real-time tracking of pathogen evolution. *Bioinformatics*. 2018;34:4121-4123. <https://doi.org/10.1093/bioinformatics/bty407>
16. Roser M, Ritchie H, Ortiz-Ospina E, Hasell J. Coronavirus pandemic (COVID-19). *Our World Data*. 2020.
17. Mercatelli D, Holding AN, Giorgi FM. Web tools to fight pandemics: the COVID-19 experience. *Brief Bioinform*. 2020:bbaa261. <https://doi.org/10.1093/bib/bbaa261>
18. Gesmann M, de Castillo D. Using the Google visualisation API with R. *R J*. 2011;3:40-44.
19. Delcher AL, Salzberg SL, Phillippy AM. Using MUMmer to identify similar regions in large sequence sets. *Curr Protoc Bioinforma*. 2003;1:10.3.1-10.3.18. <https://doi.org/10.1002/0471250953.bi1003s00>
20. Sruthi CK, Prakash MK. Statistical characteristics of amino acid covariance as possible descriptors of viral genomic complexity. *Sci Rep*. 2019;9:18410. <https://doi.org/10.1038/s41598-019-54720-y>
21. Lan J, Ge J, Yu J, et al. Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. *Nature*. 2020;581:215-220. <https://doi.org/10.1038/s41586-020-2180-5>
22. Grubaugh ND, Hanage WP, Rasmussen AL. Making sense of mutation: what D614G means for the COVID-19 pandemic remains unclear. *Cell*. 2020;182(4):794-795. <https://doi.org/10.1016/j.cell.2020.06.040>
23. Isabel S, Graña-Miraglia L, Gutierrez JM, et al. Evolutionary and structural analyses of SARS-CoV-2 D614G spike protein mutation now documented worldwide. *Sci Rep*. 2020;10:14031. <https://doi.org/10.1038/s41598-020-70827-z>
24. Becerra-Flores M, Cardozo T. SARS-CoV-2 viral spike G614 mutation exhibits higher case fatality rate. *Int J Clin Pract*. 2020;74(8):e13525. <https://doi.org/10.1111/ijcp.13525>

How to cite this article: Mercatelli D, Triboli L, Fornasari E, Ray F, Giorgi FM. *Coronapp*: A web application to annotate and monitor SARS-CoV-2 mutations. *J Med Virol*. 2021;93:3238-3245. <https://doi.org/10.1002/jmv.26678>