Published online 0 Month 0000

Nucleic Acids Research, 0000, Vol. 00, No. 00 **1–7**
doi:10.1093/nar/gkn000

CovRadar: Continuously tracking and filtering SARS-CoV-2 mutations for molecular surveillance

Alice Wittig[1,2,], Fábio Miranda[1,**], Martin Hölzer[2], Tom Altenburg[1], Jakub M. Bartoszewicz[1,2], Sebastian Beyvers[3], Marius A. Dieckmann[3], Ulrich Genske[1], Sven H. Giese[1], Melania Nowicka[1], Hugues Richard[2], Henning Schiebenhoefer[1], Anna-Juliane Schmachtenberg[1], Paul Sieben[1], Ming Tang[1,4], Julius Tembrockhaus[1], Bernhard Y. Renard[1], Stephan Fuchs[2,*]**

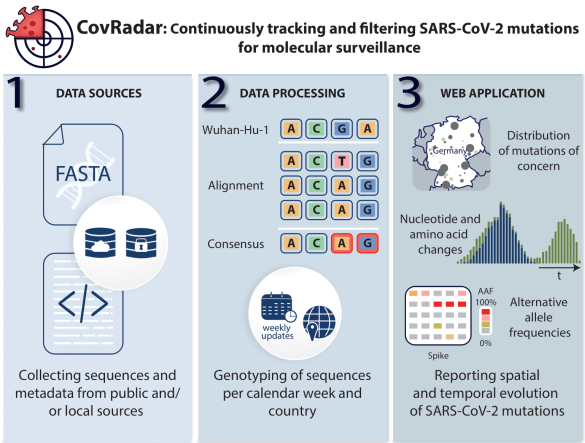
[1]Hasso Plattner Institute, Digital Engineering Faculty, University of Potsdam, Prof.-Dr.-Helmert-Straße 2 - 3, 14482 Potsdam, Germany [2]Bioinformatics Unit (MF1), Robert Koch Institute, Nordufer 20, 13353 Berlin, Germany [3]Bioinformatics & Systems Biology, Justus-Liebig-University Gießen, Heinrich-Buff-Ring 58, 35390 Gießen, Germany [4]Department of Human Genetics, Hannover Medical School, Carl-Neuberg-Str. 1, 30625 Hannover, Germany **These authors contributed equally to this work

Received Month Day, Year; Revised Month Day, Year; Accepted Month Day, Year

ABSTRACT

The SARS-CoV-2 pandemic underlined the importance of molecular surveillance to track the evolution of the virus and inform public health interventions. Fast analysis, easy visualization and convenient filtering of the latest virus sequences are essential for this purpose. However, access to computational resources, the lack of bioinformatics expertise, and the sheer volume of sequences in public databases complicate surveillance efforts. CovRadar combines an analytical pipeline and a web application designed for the molecular surveillance of the spike gene of SARS-CoV-2, an important vaccine target. The intuitive web front-end focuses on mutations rather than viral lineages and provides easy access to frequencies and spatio-temporal distributions from global sample collections. The data is regularly updated based on a scalable and reproducible analytical back-end. With this platform, we aim to give users, those with or without bioinformatics skills or sufficient computational resources, the possibility to track and explore mutational changes in the SARS-CoV-2 spike gene and to filter, download, and further analyze data that meet their questions and needs. Advanced computational users have the ability to apply the analytical pipeline and data visualization methods locally on their own data. CovRadar is freely accessible at `https://covradar.net`, source code is available at `https://gitlab.com/dacs-hpi/covradar`.

GRAPHICAL ABSTRACT



INTRODUCTION

COVID-19 is an infectious disease caused by the novel coronavirus SARS-CoV-2, which was declared a global pandemic by the World Health Organization in March 2020. Two years later, more than 434 million cases and about 5.9 million deaths have been reported worldwide (WHO, Weekly operational update on COVID-19, March 1 2022).

Molecular surveillance plays a crucial role to understand evolutionary changes and answers epidemiological questions [1]. With decreasing costs and broader access to sequencing technologies, whole-genome sequencing has proven to be an indispensable tool for molecular surveillance of SARS-CoV-2 during the pandemic [2].

*To whom correspondence should be addressed. Tel: +49 30 187544338; Email: fuchss@rki.de

© 0000 The Author(s)
This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/2.0/uk/) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

However, worldwide molecular surveillance efforts also revealed a bioinformatics bottleneck to rapidly develop specialized algorithms and tools to handle and analyze SARS-CoV-2 sequencing data and track changes in the pandemic faster [3].

In March 2022, the COVID-19 Data Portal hosted by EMBL-EBI [4] includes more than 4 million genome sequences and the international SARS-CoV-2 database GISAID [5, 6, 7] more than 9 million genome sequences, resulting in a 94 GB and a 250 GB multi-FASTA file, respectively. The sheer amount of data makes the analysis and search for known, new, and emerging mutations in the viral genome even more complicated and requires not only bioinformatics expertise but also significant computational resources; at the same time, comprehensive analysis of these data is essential to detect new viral variants, determine fixation rates, and ensure the efficacy of vaccines and diagnostic tests. Various tools were therefore developed in a short amount of time to tackle such challenges and to process the data in different ways while maintaining easy access to those with no computational background.

We performed a literature search and found 17 online tools available in the field of molecular surveillance of SARS-CoV-2 (Supplementary Table S1). For example, Nextstrain (nextstrain.org) [8] provides a platform for real-time tracking of SARS-CoV-2 evolution but needs to significantly down-sample sequences for calculations and visualizations. Other commonly used platforms such as outbreak.info and covariants.org also provide insights into mutations but group them based on already defined SARS-CoV-2 lineages [9] or lack options to examine detailed spatial distributions when the metadata contains such information. With Cov-Spektrum (https://cov-spectrum.org) [10], it is possible to monitor variants and mutations, but the tool lacks options to directly compare different countries, time periods, or hosts. We continue to see an urgent need to make SARS-CoV-2 sequence information more readily available to virologists and epidemiologists on a daily basis, especially in a public health context.

Here we present CovRadar, which was developed to address the aforementioned challenges. CovRadar consists of an analytical pipeline and an intuitive web application hosted on a powerful and scalable Kubernetes cluster (Fig. 1). The analytical workflow can process millions of genome sequences and can also combine different data sources. CovRadar functions primarily at the mutational level, allowing for early observations even before a new SARS-CoV-2 variant is assigned an official lineage name. Researchers can explore vast amounts of already pre-processed genome data, select specific subsets in space and time, and export their analyses for further investigations. Both, the pipeline and the web application can be installed and run on a local system to include access-restricted sequences and additional metadata. The data sources underlying CovRadar's visualizations are updated regularly and therefore allow tracking of SARS-CoV-2 mutations in the spike gene in near real-time, so that threatening changes can be detected quickly and local health authorities can be informed and respond.

MATERIALS AND METHODS

Overview

The web application (covradar.net) is implemented using a Python-based web framework, Flask (github.com/pallets/flask), that comprises two Dash applications (https://github.com/plotly/dash), each connected to a MariaDB database. One database includes the results of the analytical pipeline (Fig. 1 and Fig. S1) for sequences of the COVID-19 Data Portal hosted by EMBL-EBI (covid19dataportal.org) and is updated bi-weekly, the other database includes the sequences of the last four weeks and is updated every other day. All back-end functionalities of the web application are written in Python. Images and tables are produced via Plotly.js (https://plot.ly). By having both the back- and front-end written in Python, we can more easily maintain the code and integrate further interactive user interfaces while keeping the Python syntax for the back-end logic and algorithms, instead of using PHP or JavaScript in the front-end. The design principles of CovRadar's web application are focused on intuitive overviews that allow users to customize the visualized data based on their needs and to further download the filtered data sets for *post-hoc* specialized analyses. All elements such as plots and tables are interactive and can be directly modified and downloaded in PNG format and raw data format (CSV) for further processing. The web application runs on a Kubernetes cluster hosted by de.NBI. A local version can be started with the integrated Flask server or with a provided Docker container. Detailed instructions for installation can be found on https://gitlab.com/dacs-hpi/covradar.

Position converter

To ease coordinate based queries, we implemented a converter between genome and protein sequences (accessible from the right sidebar of the website). The user can easily convert a range of genomic positions into (codon) positions within the spike gene or amino acids multiple sequence alignment (MSA) positions. For example, the amino acid change D614G corresponds to nucleotide positions $1840-1842$ in the spike gene, which represents genomic coordinates $23402-23404$ on the reference (Wuhan-Hu-1). The converter script can be also used as a stand-alone tool supporting all genes encoded by the SARS-CoV-2 genome (gitlab.com/dacs-hpi/covradar/-/tree/master/shared/residue_converter).

Geographic distribution map of mutation frequencies

An interactive map shows the presence and frequency of selected amino acid changes for a given time period. For the geographic distribution, we use the highest resolution metadata available, which are postal codes for the German data (obtained via DESH, the German Electronic Sequence

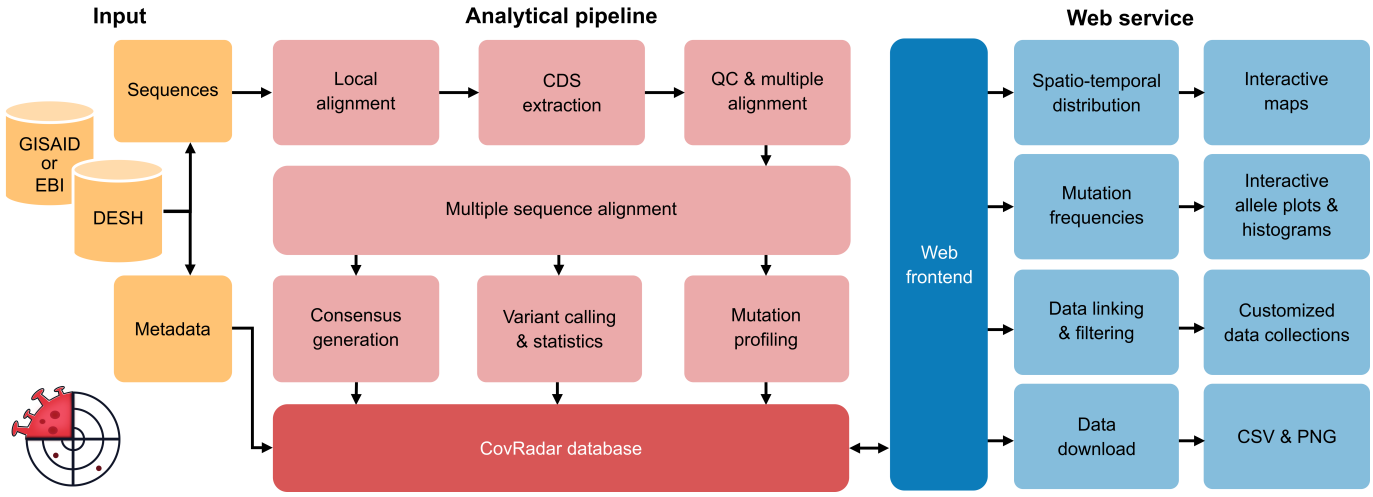


Figure 1. Simplified overview of Covradar's input, analytical pipeline, and web service. The analytical pipeline is described in detail in the Supplementary Methods (Fig. S1). DESH – The German Electronic Sequence Data Hub which we combine with data from e.g. EMBL-EBI's COVID-19 Data Hub for higher national resolution.

Data Hub) and countries otherwise (see Supplementary Methods). Importantly, the sampling locations for the German data are based on the postal codes of the primary diagnostic laboratories where a PCR test was performed to select a sample for sequencing. Therefore, the mutation frequencies shown for the Germany map do not necessarily correspond with the real distribution of SARS-CoV-2 mutations in Germany and do not allow direct conclusions about the mutation distribution in an area. Nevertheless, by integrating our national data source, we achieve a higher geographic resolution in comparison to what is possible only based on GISAID or EMBL-EBI data and for other countries. The user can select between multiple visualization modes. The mode "n-th most frequent mutations per location" shows for n=1 the most frequent mutation per location, for n=2 the second most frequent mutation, and so on. The map can also show the frequencies of selected mutations per location. The "mutation proportion" mode can be used to check how many sequences have at least one of the selected mutations. Finally, the user can also examine the increase per location determined by linear regression within the selected time interval.

Clicking on a location within the map further shows information for that location in two plots: 1) A bar chart showing the number of sequences (mode "frequency") and slopes (mode "increase"), respectively, as well as 2) the time course of mutation frequencies of a selected mutation and for a certain geographic location using linear regression.

Alternative allele frequency plot

We implemented a heatmap-inspired alternative allele frequency (AAF) plot to visualize the frequencies of mutated sites (substitutions, INDELs) in the spike gene (Fig. 2). Each block represents a nucleotide in the MSA of extracted spike sequences. The coordinates on the y-axis correspond to the positions in the MSA. The colors represent the AAF of the selected sequences in comparison to a reference sequence (default: Wuhan-Hu-1, NC_045512.2). The mouse-over tooltips show the user the position references as well as the AAF and the coverage. Coverage describes how many sequences are used to retrieve the AAF at this position. Importantly, the analysis of the consensus sequences is not limited to the Wuhan-Hu-1 reference spike but it can also be compared to a large selection of consensus sequences that are provided per calendar week and country (see also Fig. S2). By that, the user can, for example, compare the current mutational pattern in Germany against the consensus sequence from another country in the same (or different) time period. The filters allow only specific SARS-CoV-2 lineages, hosts, time intervals, and countries to be selected. For better interpretation of the selected data, the number of sequences in the consensus sequence and in the filtered data set are shown above the plot. By default, the region encoding the receptor binding domain (RBD) of the spike gene is highlighted, but additional domains can also be activated, as well as any mutation positions shown in the geographic distribution map and custom amino acid/codon positions. This can be used to highlight specific locations when the plot is exported as a PNG. Further details about how AAF are calculated can be found in the Supplementary Methods.

Mutation distribution plots

The nucleotide distribution plot shows the frequency of different nucleotides at a given spike gene for each calendar week and country. Accordingly, an amino acid distribution plot highlight the frequency of different amino acids encoded by the respective codon (Fig. 3).

Front-end specifications

The CovRadar web service (`covradar.net`) uses the storage and compute resources of a dedicated Kubernetes (`https://kubernetes.io/de/`) cluster provided by the German Network for Bioinformatics Infrastructure (`de.NBI`). A MariaDB database, Redis caches, and web services are deployed in a redundant setup that allows for constant availability of the provided services. If necessary, the complete

infrastructure setup can be easily scaled up on demand both in terms of storage volume and computing resources. All popular browsers are supported (see Table S2 for details). The database is hosted on local SSD storage to boost read performance. Database replication to ensure persistence is not needed as the database is working in a read-only-mode after the initial creation, and can be easily re-created in case of data loss. The whole setup is designed to provide a robust, scalable and performant infrastructure to run the CovRadar service.

RESULTS

CovRadar adds value via mutation-focused spike surveillance

During the pandemic, several tools were developed to analyze and visualize genomic and epidemiologic data to help track evolution and spread of the virus. Here, we collected web-based tools with a focus on SARS-CoV-2 data analysis and compared them to CovRadar's features (Table S1). We want to briefly mention that several of the selected tools also provide additional functionalities, such as phylogenies and protein structures, that were not highlighted in CovRadar. We have developed CovRadar not to compete with existing platforms, but to add real value by focusing on individual mutations of the spike gene, which encodes an important target for vaccines. Thus, we focus here on a comparison of CovRadar's features with the same features of the selected tools.

We collected 17 resources with a focus on their availability and SARS-CoV-2 relevance using Aviator [11], a tool for monitoring web services based on literature, OverCOVID (`http://bis.zju.edu.cn/overcovid/resource.html`), and our experience in daily work in SARS-CoV-2 surveillance. We excluded all non-molecular surveillance tools and non-working tools (although still online), resulting in this selection: GESS [12], Global evolution of SARS-CoV-2 [13], EMBL-EBI COVID-19 Data Portal [4], GISAID CoVSurver [6], CovidPhy [14], Coronavirus3D, NCBI, PathogenWatch, Nextstrain [8], CoV-GLUE [15], Outbreak, coVariants, CoV-Spectrum [10], COG-UK Mutation Explorer, COVID-19 variant dashboard, CoVerage, and Monitor of SARS-CoV-2 variants (Table S1). Most of these tools focus on information about lineages and their distribution over time, while only a few also provide distributions and spatial information for single mutations. In most cases, it is not possible to display information for different time periods on a map comparing multiple mutations. CoV-Spectrum is the only tool that provides a comparison with other variants and mutations, but it was not possible to compare different countries, calendar weeks or to select hosts. No tool offered alternative allele frequencies. In addition, no tool offered a built-in position converter, which is particularly useful when searching for mutations from the literature in an MSA. Other features are listed and compared in Table S1.

Mutation frequency analyses on national level shows increasing frequencies of characteristic Omicron mutations within the last two weeks of December 2021

In this example, we investigate recent mutation dynamics within the spike gene with a focus on a specific country and time period. We use CovRadar's filter options, allele frequency and mutation distribution plots, to show the rise of characteristic mutations in the spike gene in a short time frame and for the end of the year 2021 in Germany (Fig. 2 and 3). By taking delays in sampling, sequencing, and data submission into account, we restrict the data set to sequences from Germany of the last two calendar weeks in 2021 (CW51+52) and compare them to the consensus sequence from CW50. Thus, we only see the differences between consensus sequences specific for this time period and German sequences and not in comparison to Wuhan-Hu-1. Based on these filters, we can compare the AAF of 19,890 sequences (CW51+52) against the consensus based on 14,282 sequences (CW50) (Fig. 3).

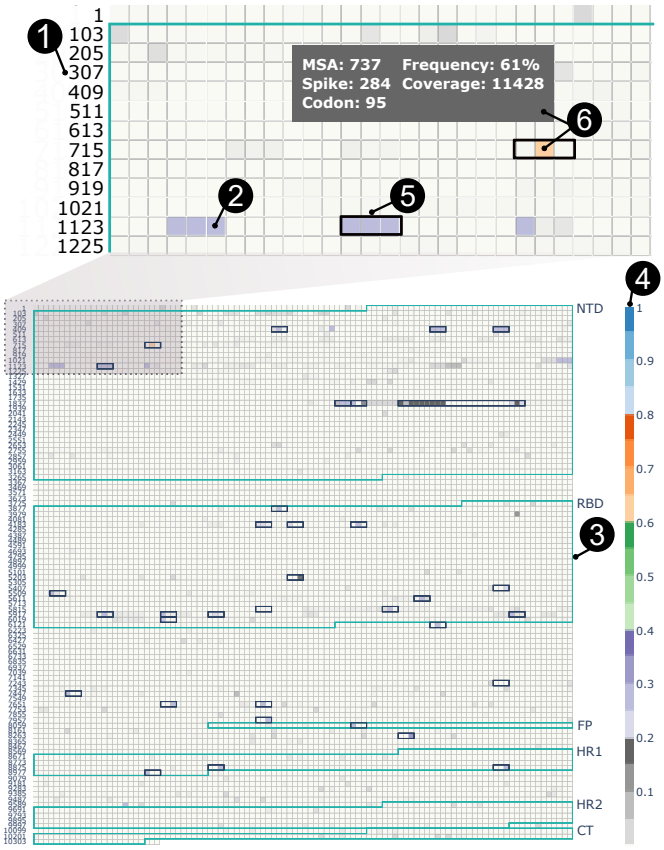


Figure 2. Alternative allele frequency (AAF) plot of German sequences from calendar weeks (CW) 51 and 52 of 2021 compared to the German consensus sequence of CW50. Coordinates on the left correspond to the MSA positions **(1)** and can be easily converted via the position converter or via mouse over. Each block represents a nucleotide position in the MSA of spike gene sequences **(2)**. Spike domains are highlighted with blue frames **(3)** and labeled NTD: N-terminal Domain, RBD: Receptor Binding Domain, FP: fusion peptide, HR1 and 2: heptapeptide repeat sequence 1 and 2, CT: cytoplasmic domain fusion. The color of the block represents the AAF **(4)**. The plot shows increased frequencies for nucleotide positions common for Omicron (mutations with $>75\,\%$ frequency based on `outbreak.info`, black frames **(5)**). The highest AAF in this comparison ($\sim$61 %) belongs to codon 95 (orange box, **(6)**). Shown data is based on the DESH data set.

Sites characteristic of Omicron (mutations with $>75\%$ frequency in Omicron based on `outbreak.info`) show a higher AAF and are distributed among the different spike domains such as the Receptor Binding Domain (RBD) and the N-terminal domain (NTD) (see Fig. 2). The highest AAF

at position 284 of the spike gene is within the NTD of the respective gene product (position 95 at the protein level). Taking a closer look on alternate alleles at this position (using the mutation distribution plot), we see that C has been replaced by T at position 284 which results in the T95I mutation profile on protein level (Fig. 3). The T95I change is characteristic for Omicron but has been previously described also for the Variant of Interest Mu and the Variant under Monitoring Iota and has been reported in vaccine breakthrough infections [16]. Furthermore, a pathogenicity analysis predicted T95I, among other mutations, to be deleterious in Omicron and sub-lineages [17]. We downloaded the complete time series data for C284T from the mutation distribution plot via the web app to perform an additional trend analysis. We compared the proportion of the alternative nucleotide T at position 284 between CWs using a Fisher exact test (adjusted p-value < 0.01, Table S3, see [18] for details on the method). We found that in two time periods (CW21-34 and CW 47-52), the proportion of samples with the alternative nucleotide T at position 284 increased significantly between consecutive weeks (Fig. 3). Based on these observations and analyses, we can expect T95I to become consensus, which is indeed the case in CW51 and subsequent weeks.

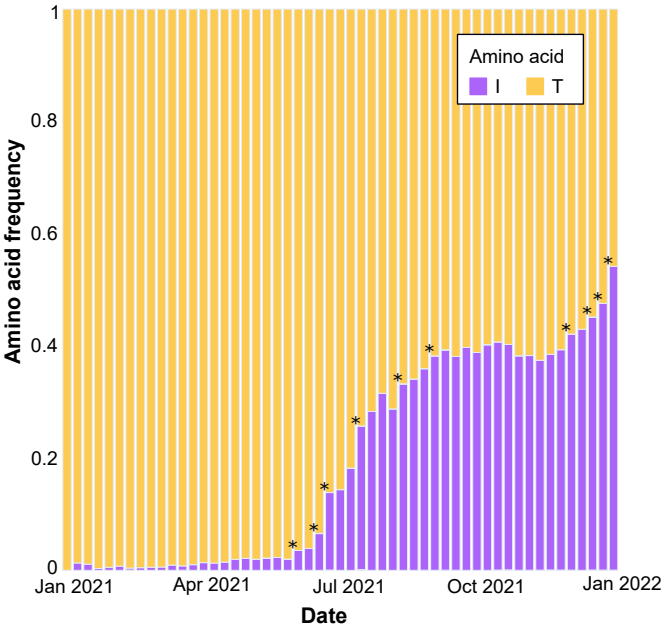


Figure 3. Frequency of amino acids at codon position 95 over calendar weeks (CW) for German sequences. The reference amino acid is threonine (T, yellow) and we can observe an increase in isoleucine (I, purple). Based on the underlying time series data downloaded from CovRadar, we calculated a significant increase of T95I in two time periods (CW21–34 and CW 47–52). Significant increases between consecutive CWs are marked by an asterisk and were calculated as described before [18].

Early emergence of spike change L452R in May 2021 in Germany

Using CovRadar, users can easily track spatio-temporal occurences of spike mutations, for example, linked to Variants of Concern (VOC). By default, we obtain information about characteristic VOC mutations from `outbreak.info` that can be selected by the user and then highlighted in a map. In this retrospective analysis, we investigate if characteristic Delta mutations were already evident in Germany in May 2021, a time where Alpha was still dominant. We select May 6 - 20, 2021 as the time frame and focus on the DESH data set with available postal codes. According to a weekly situation report by the Robert Koch Institute, Delta became dominant in Germany in CW25 (June, 2021) [19]. First, we selected N501Y which is strongly associated with VOC Alpha, Beta, and Gamma and E484K, known from Beta and Gamma. Both changes are only rarely observed in Delta ($<0.1\%$, `outbreak.info`). Second, we also selected L452R and T478K, which are both strongly represented in Delta, but only with low frequencies $<0.1\%$ in Alpha, Beta, Gamma. The L452R change has been shown to increase infectivity *in vitro* and decrease neutralization by sera from COVID-19 patients and vaccinated individuals [20, 21] and has also been detected in other lineages such as A.27 which rose during the early months in 2021 in Germany [22]. The T478K change is commonly known from Delta and Omicron and has been predicted to affect the spike/ACE2 interaction [23]. Almost 100% of the sequences (21,128 / 21,156) that fall into the selected time frame (May 6 - 20, 2021) contain at least one of the four selected mutations (L452R, T478K, N501Y, E484K; Supplementary Fig. S3). We found that in May 2021, N501Y was dominant (Supplementary Fig. S3) but we can already observe a positive trend for other mutations such as L452R (Fig. 4) in some German regions.

To further investigate in which geographic location the first significant rise in L452R can be observed, we downloaded the selected time series data per location. We tested for significant differences in the proportion within the single postal code areas compared between calendar weeks (Fisher exact test, adjusted p-value <0.01, Table S4, see [18] for details). We excluded data from February 2021 to rule out fluctuation effects due to increased sequencing capacity starting from January 25, 2021 in Germany. The first significant increase, excluding February, was sequenced in CW16 2021 (April 19–25 2021) for postal code 50933 (Cologne, North Rhine-Westphalia). In addition, we found that L452R significantly increased in CW31 with an 18 fold change in the postal code area 20095 (Hamburg). These examples also illustrate the importance of granular geographic data to track specific mutations and take action to contain new variants as quickly as possible. Please note that all sampling locations presented here are based on the primary diagnostic labs (performing the PCR test) and not the location of the patients due to data protection regulations.

CONCLUSION

Molecular surveillance plays a crucial role in tracking the spread and evolution of SARS-CoV-2. Worldwide, millions of genome sequences have been generated and uploaded to databases such as GISAID and the COVID-19 Data Portal. While such data sources are treasure troves to answer various questions, the sheer amount of sequencing data directly leads to algorithmic and computational challenges, especially for non-bioinformaticians. Many web-based tools have been developed to provide fast access to already pre-processed data. However, these tools focus primarily on information about the SARS-CoV-2 lineage and Variants of Concern whose definition underlies a delay, while emerging Mutations of

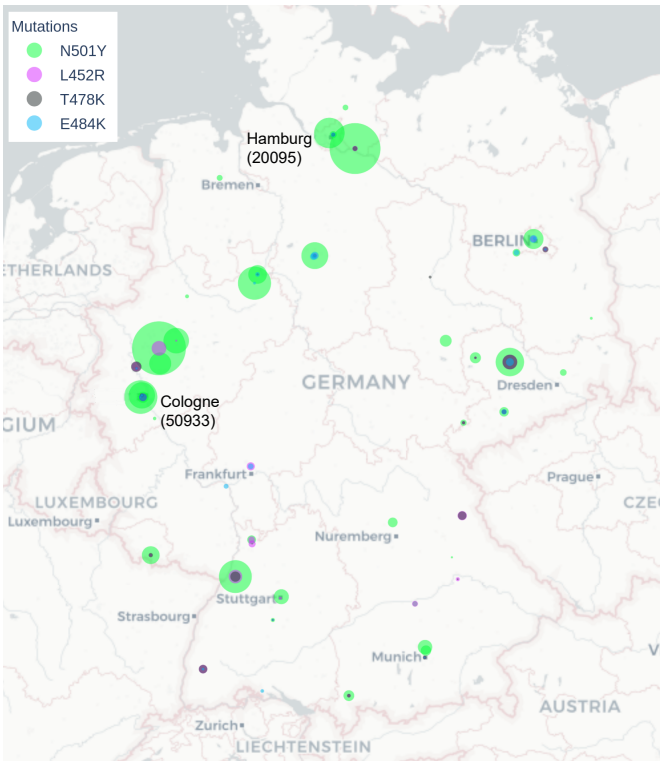


Figure 4. The geographic map shows the slope (circle size) of positive trends for the selected spike mutations L452R, T478K, E484K, and N501Y from May 6, 2021 to May 20, 2021. While N501Y is dominant throughout sampling locations (primary diagnostic labs) in Germany, we also observe a rise in other mutations such as L452R in certain areas.

Concern are often neglected. Thus, we present CovRadar, a web service to enhance molecular surveillance of the spike gene with a focus on the mutation level and customizable spatial and time analyses. We believe that CovRadar will aid researchers around the world to get better and faster access to SARS-CoV-2 mutation profiles and metadata to support the ongoing fight against the COVID-19 pandemic on the molecular surveillance level.

AVAILABILITY

CovRadar is freely accessible at `https://covradar.net`, its open-source code is available at `https://gitlab.com/dacs-hpi/covradar`.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR online.

ACKNOWLEDGEMENTS

Many thanks to Elizabeth Yuu for proofreading the manuscript. We are very grateful to the GISAID Initiative, the COVID-19 Data Portal hosted by EMBL-EBI, the German Electronic Sequence Data Hub (DESH, `https://github.com/robert-koch-institut/SARS-CoV-2-Sequenzdaten_aus_Deutschland`) and all data contributors, i.e. the authors from the originating laboratories responsible for obtaining the specimens and the submitting laboratories where genetic sequence data were generated and shared and on which this research is based.

Funding This work was supported by the European Centers for Disease Control [grant number ECDC GRANT/2021/008 ECD.12222] to SF, the Bundesministerium für Bildung und Forschung [grant numbers 031L0175B, 01KI1905D] to BYR as well as BMWK DAKI [grunt number 01MK21009E] to BYR and SF, and the German Network for Bioinformatics Infrastructure; de.NBI-cloud [grant numbers 031A537B, 031A533A, 031A538A, 031A533B, 031A535A, 031A537C, 031A534A, 031A532B].

Conflict of interest statement. None declared.

REFERENCES

1. Lee W. Riley and Ronald E. Blanton. Advances in molecular epidemiology of infectious diseases: Definitions, approaches, and scope of the field. *Microbiol. Spectr.*, 6(6), 2018. [PubMed:30387413] [PubMed Central:PMC6343655] [doi:10.1128/microbiolspec.ame-0001-2018].

2. Lucy van Dorp, Charlotte J Houldcroft, Damien Richard, and François Balloux. COVID-19, the first pandemic in the post-genomic era. *Curr. Opin. Virol.*, 50:40–48, 2021. [PubMed:34352474] [PubMed Central:PMC8275481] [doi:10.1016/j.coviro.2021.07.002].

3. Emma B. Hodcroft, Nicola De Maio, Rob Lanfear, Duncan R. MacCannell, Bui Quang Minh, Heiko A. Schmidt, Alexandros Stamatakis, Nick Goldman, and Christophe Dessimoz. Want to track pandemic variants faster? fix the bioinformatics bottleneck. *Nature*, 591(7848):30–33, 2021. [PubMed:33649511] [doi:10.1038/d41586-021-00525-x].

4. Peter W Harrison, Rodrigo Lopez, Nadim Rahman, Stefan Gutnick Allen, Raheela Aslam, Nicola Buso, Carla Cummins, Yasmin Fathy, Eloy Felix, Mihai Glont, et al. The COVID-19 Data Portal: accelerating SARS-CoV-2 and COVID-19 research through rapid open access data sharing. *Nucleic Acids Res.*, 49(W1):W619–W623, 2021. [PubMed:34048576] [PubMed Central:PMC8218199] [doi:10.1093/nar/gkab417].

5. Stefan Elbe and Gemma Buckland-Merrett. Data, disease and diplomacy: GISAID's innovative contribution to global health. *Glob. Chall.*, 1(1):33–46, 2017. [PubMed:31565258] [PubMed Central:PMC6607375] [doi:10.1002/gch2.1018].

6. Yuelong Shu and John McCauley. GISAID: Global initiative on sharing all influenza data – from vision to reality. *Eurosurveillance*, 22(13), 2017. [PubMed:28382917] [PubMed Central:PMC5388101] [doi:10.2807/1560-7917.es.2017.22.13.30494].

7. Shruti Khare, Céline Gurry, Lucas Freitas, Mark B Schultz, Gunter Bach, Amadou Diallo, Nancy Akite, Joses Ho, Raphael TC Lee, Winston Yeo, et al. GISAID's Role in Pandemic Response. *China CDC Wkly.*, 3(49):1049, 2021. [PubMed:34934514] [PubMed Central:PMC8668406] [doi:10.46234/ccdcw2021.255].

8. James Hadfield, Colin Megill, Sidney M Bell, John Huddleston, Barney Potter, Charlton Callender, Pavel Sagulenko, Trevor Bedford, and Richard A Neher. Nextstrain: real-time tracking of pathogen evolution. *Bioinformatics*, 34(23):4121–4123, 2018. [PubMed:29790939] [PubMed Central:PMC6247931] [doi:10.1093/bioinformatics/bty407].

9. Áine O'Toole, Emily Scher, Anthony Underwood, Ben Jackson, Verity Hill, John T McCrone, Rachel Colquhoun, Chris Ruis, Khalil Abu-Dahab, Ben Taylor, et al. Assignment of Epidemiological Lineages in an Emerging Pandemic Using the Pangolin Tool. *Virus Evol.*, 2021. [PubMed:34527285] [PubMed Central:PMC8344591] [doi:10.1093/ve/veab064].

10. Chaoran Chen, Sarah Nadeau, Michael Yared, Philippe Voinov, Ning Xie, Cornelius Roemer, and Tanja Stadler. CoV-spectrum: analysis of globally shared SARS-CoV-2 data to identify and characterize new variants. *Bioinformatics*, 38(6):1735–1737, 2021. [PubMed:34954792] [PubMed Central:PMC8896605] [doi:10.1093/bioinformatics/btab856].

11. Tobias Fehlmann, Fabian Kern, Pascal Hirsch, Robin Steinhaus, Dominik Seelow, and Andreas Keller. Aviator: a web service for monitoring the availability of web services. *Nucleic Acids Res.*, 49(W1):W46–W51, 2021. [PubMed:34038559] [PubMed Central:PMC8262725] [doi:10.1093/nar/gkab396].

12. Shuyi Fang, Kailing Li, Jikui Shen, Sheng Liu, Juli Liu, Lei Yang, Chang-Deng Hu, and Jun Wan. GESS: a database of global evaluation of SARS-CoV-2/hCoV-19 sequences. *Nucleic Acids Res.*, 49(D1):D706–D714, 2021. [PubMed:33045727] [PubMed Central:PMC7778918] [doi:10.1093/nar/gkaa808].

13. Xiaolu Tang, Ruochen Ying, Xinmin Yao, Guanghao Li, Changcheng Wu, Yiyuli Tang, Zhida Li, Bishan Kuang, Feng Wu, Changsheng Chi, et al. Evolutionary analysis and lineage designation of SARS-CoV-2 genomes. *Sci. Bull.*, 66(22):2297–2311, 2021. [PubMed:33585048] [PubMed Central:PMC7864783] [doi:10.1016/j.scib.2021.02.012].

14. Xabier Bello, Jacobo Pardo-Seco, Alberto Gómez-Carballa, Hansi Weissensteiner, Federico Martinón-Torres, and Antonio Salas. CovidPhy: A tool for phylogeographic analysis of SARS-CoV-2 variation. *Environ. Res.*, 204:111909, 2022. [PubMed:34419470] [PubMed Central:PMC8376833] [doi:10.1016/j.envres.2021.111909].

15. Joshua Singer, Robert Gifford, Matthew Cotten, and David Robertson. CoV-GLUE: a web application for tracking SARS-CoV-2 genomic variation. *Preprints*, 2020. [doi:10.20944/preprints202006.0225.v1].

16. Ezgi Hacisuleyman, Caryn Hale, Yuhki Saito, Nathalie E. Blachere, Marissa Bergh, Erin G. Conlon, Dennis J. Schaefer-Babajew, Justin DaSilva, Frauke Muecksch, Christian Gaebler, et al. Vaccine Breakthrough Infections with SARS-CoV-2 Variants. *N. Engl. J. Med.*, 384(23):2212–2218, 2021. [PubMed:33882219] [PubMed Central:PMC8117968] [doi:10.1056/NEJMoa2105000].

17. Suresh Kumar, Kalimuthu Karuppanan, and Gunasekaran Subramaniam. Omicron (BA.1) and sub-variants (BA.1, BA.2 and BA.3) of SARS-CoV-2 spike infectivity and pathogenicity: A comparative sequence and structural-based computational assessment. *bioRxiv*, 2022. [doi:10.1101/2022.02.11.480029].

18. Sébastien Calvignac-Spencer, Matthias Budt, Matthew Huska, Hugues Richard, Luca Leipold, Linus Grabenhenrich, Torsten Semmler, Max von Kleist, Stefan Kröger, Thorsten Wolff, and Martin Hölzer. Rise and Fall of SARS-CoV-2 Lineage A.27 in Germany. *Viruses*, 13(8), 2021. [PubMed:34452356] [PubMed Central:PMC8402818] [doi:10.3390/v13081491].

19. Djin-Ye Oh, Martin Hölzer, Sofia Paraskevopoulou, Maria Trofimova, Felix Hartkopf, Matthias Budt, Marianne Wedde, Hugues Richard, Berit Haldemann, Teresa Domaszewska, et al. Advancing precision vaccinology by molecular and genomic surveillance of SARS-CoV-2 in germany. *Under review in Clinical Infectious Diseases*, 2022.

20. Qianqian Li, Jiajing Wu, Jianhui Nie, Li Zhang, Huan Hao, Shuo Liu, Chenyan Zhao, Qi Zhang, Huan Liu, Lingling Nie, et al. The impact of mutations in SARS-CoV-2 spike on viral infectivity and antigenicity. *Cell*, 182(5):1284–1294, 2020. [PubMed:32730807] [PubMed Central:PMC7366990] [doi:10.1016/j.cell.2020.07.012].

21. Xianding Deng, Miguel A Garcia-Knight, Mir M Khalid, Venice Servellita, Candace Wang, Mary Kate Morris, Alicia Sotomayor-González, Dustin R Glasner, Kevin R Reyes, Amelia S Gliwa, et al. Transmission, infectivity, and neutralization of a spike L452R SARS-CoV-2 variant. *Cell*, 184(13):3426–3437, 2021. [PubMed:33991487] [PubMed Central:PMC8057738] [doi:10.1016/j.cell.2021.04.025].

22. Tamara Kaleta, Lisa Kern, Samuel Hong, Martin Hölzer, Georg Kochs, Julius Beer, Daniel Schnepf, Martin Schwemmle, Nena Bollen, Philipp Kolb, et al. Antibody escape and global spread of SARS-CoV-2 lineage A.27. *Nat. Commun.*, 13(1152), 2022. [PubMed:35241661] [doi:10.1038/s41467-022-28766-y].

23. Simone Di Giacomo, Daniele Mercatelli, Amir Rakhimov, and Federico M Giorgi. Preliminary report on severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) spike mutation t478k. *J. Med. Virol.*, 93(9):5638–5643, 2021. [PubMed:33951211] [PubMed Central:PMC8242375] [doi:10.1002/jmv.27062].