In silico genomic surveillance by CoVerage predicts and characterizes SARS-CoV-2 Variants of Interest

K. Norwood*[,1,2], Z.-L. Deng*[,1], S. Reimering*[1], G. Robertson[1], M.-H. Foroughmand-Araabi[1,2], S. Goliaei[1,2], M. Hölzer[3] and A. C. McHardy[1,2,4]

[1]Computational Biology of Infection Research, Helmholtz Centre for Infection Research, Braunschweig, Germany

[2]Braunschweig Integrated Centre of Systems Biology (BRICS), Technische Universität Braunschweig, Braunschweig, Germany

[3]Genome Competence Center (MF1), Robert Koch Institute, Berlin, Germany

[4]Deutsches Zentrum für Infektionsforschung (DZIF), Braunschweig, Germany

*These authors contributed equally to this work

Correspondence to A.C. McHardy: amc14@helmholtz-hzi.de

Abstract

Rapidly evolving viral pathogens such as SARS-CoV-2 continuously accumulate amino acid changes, some of which affect transmissibility, virulence or improve the virus' ability to escape host immunity. Since the beginning of the pandemic and establishment of SARS-CoV-2 as a human pathogen, multiple lineages with concerning phenotypic alterations, so called Variants of Concern (VOCs), have emerged and risen to predominance. To optimize public health management and to ensure the continued efficacy of vaccines, the early detection of such variants of interest is essential. Therefore, large-scale viral genomic surveillance programs have been initiated worldwide, with data being deposited in public repositories in a timely manner. However, technologies for their continuous interpretation are currently lacking. Here, we describe the CoVerage system (www.sarscoverage.org) for viral genomic surveillance, which continuously predicts and characterizes novel and emerging potential Variants of Interest (pVOIs) together with their antigenic and evolutionary alterations. Using the establishment of Omicron and its current sublineages as an example, we demonstrate how CoVerage can be used to quickly identify and characterize such variants. CoVerage can facilitate the timely identification and assessment of future SARS-CoV-2 Variants of Concern.

Introduction

In early 2020, infections with a previously unknown coronavirus of probable zoonotic origin in Wuhan, China were first reported[1]. The virus, named SARS-CoV-2, rapidly spread across the globe, causing over 675 million infections and 6.8 million deaths as of March 2023[2]. SARS-Cov-2 has a single-stranded, positive-sense RNA genome with substantial capacity to mutate, reflected in its strain-level diversity of circulating viral lineages and rapid evolution[3,4]. This led to the emergence of several Variants of Concern (VOCs), as designated by the WHO, with altered phenotypes in transmissibility, virulence or antigenicity, causing large waves of new infections or reinfections[5,6,7,8]. One of the most recent VOCs, Omicron or B.1.1.529 in the Pangolin notation[9], exhibits both increased transmissibility and antigenic drift, allowing escape from immunity acquired by prior infection or vaccination[10,11] and quickly replaced almost all other lineages worldwide.

To monitor SARS-CoV-2 evolution and adaptation as well as to enable the timely identification of new VOCs, many countries have implemented large-scale viral genomic surveillance programs, leading to the generation of unprecedented amounts of sequence data. In the fall of 2023, more than 16 million sequences were available in the GISAID database[12]. Analyzing this data requires scalable computational methods for processing. Currently, there are web-based visualization platforms such as NextStrain, CoVariants, CovidCG, outbreak.info, CoVRadar, COVID-19 Viral Genome Analysis Pipeline[13-18], among others. These platforms offer various analyses based on publicly available sequencing data, further supporting scientists, public health officials, and the general public in making sense of these highly complex data. However, there is a delay in the availability of sequences between the time of collection, submission to databases such as GISAID, and distribution. Moreover, none of these resources directly identify potential Variants of Interest (pVOIs), i.e., lineages that may have a selective advantage to spread in the host population, from the available data.

Here, we describe CoVerage (https://sarscoverage.org), a web-based analytics platform that continuously monitors the genetic and antigenic evolution of human SARS-CoV-2 viruses. The platform implements methods that continuously search for pVOIs and predicts their antigenic alterations relative to the original Wuhan strain, using publicly available viral surveillance data in a fully automated, timely fashion. Thus, CoVerage can support the early detection of future VOCs. We demonstrate the application of the framework for the early detection of VOCs using the

example of the Omicron variant that appeared in November of 2021, as well as for finding circulating pVOIs in the first quarter of 2023.

Results

CoVerage site implementation

The CoVerage analytical workflow comprises several stages: (1) input data acquisition and filtration, (2) computational sequence data analyses, (3) creation of template pages based on a bootstrap framework, and (4) visualization in the browser using GitHub Pages (**Fig. 1**). First, both genomic sequences and the corresponding metadata file are used as inputs for the CoVerage computational analytics pipeline. For this purpose, genomic sequences and the sequence metadata file are downloaded with daily updates from GISAID[19]. Optionally, German sequence data is downloaded from Zenodo, where it is made available by the Robert Koch Institute (RKI) (https://zenodo.org/record/8334829), prior to its submission to GISAID. Case numbers are obtained from the WHO Coronavirus (COVID-19) data repository[20]. The computational workflow performs the analysis of SARS-CoV-2 lineage dynamics by country to identify emerging lineages that may possess a selective advantage, analysis of allele dynamics of the SARS-CoV-2 major surface protein to identify alleles with amino acid changes that may provide a selective advantage, and the *in silico* assessment of antigenic lineage alterations. Antigenic alteration and the lineage dynamics analysis require the GISAID metadata as input, while the spike protein allele dynamics requires both the genomic sequences and associated metadata to infer the coding sequences. Each of these analyses and data downloads are run independently twice per week, and the results are updated on the web-server accordingly.

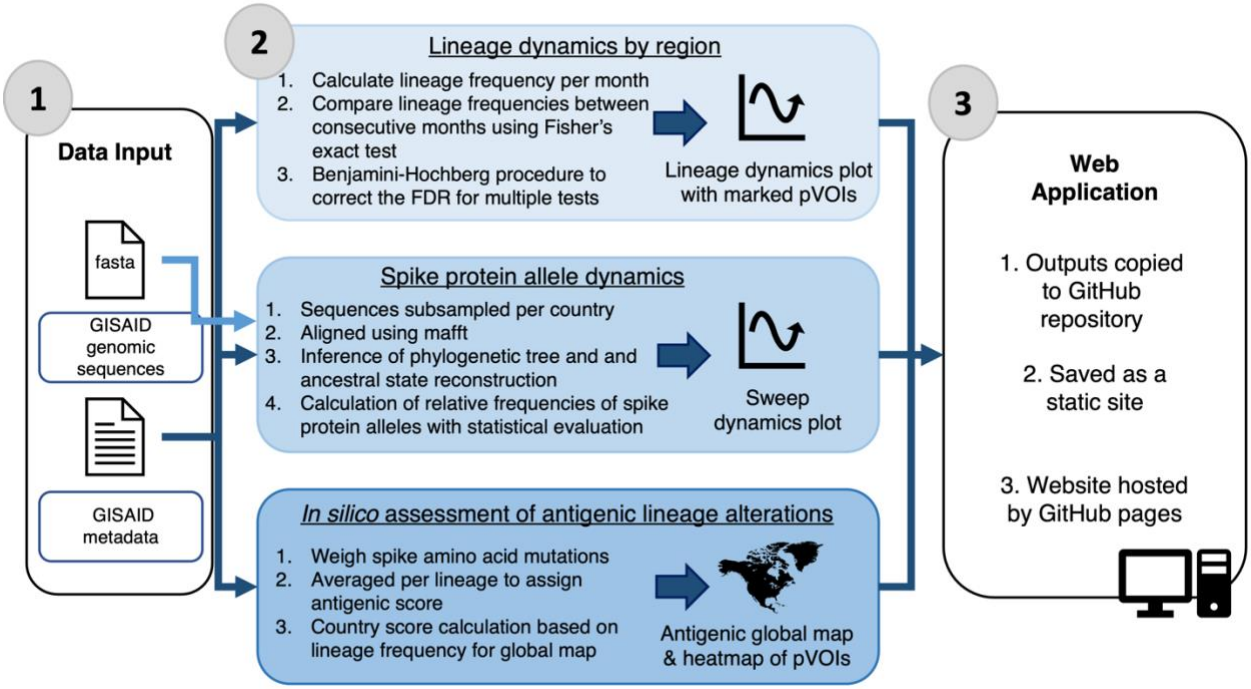

Fig. 1: Continuous viral genome analytics provided by the CoVerage platform (sarscoverage.org). Data input includes GISAID genomic sequences and metadata for SARS-CoV-2 isolates from GISAID. These inputs are then fed through three computational workflows, whose outputs are saved to individual GitHub repositories for each analysis. The results are then copied to the main CoVerage GitHub repository, where the individual pages are saved as static sites, which are hosted by GitHub pages. Workflows are run twice a week and result pages are updated accordingly.

Lineage dynamics by region

For the analysis of lineage dynamics and to identify pVOIs, i.e. lineages that may have a selective advantage to spread in the host population, we adapted a technique that we developed for recommending vaccine strain updates for the seasonal influenza vaccine[21] (https://github.com/hzi-bifo/SDplots). There are three methodologies available for identifying pVOI lineages: the standard method identifies lineages increasing significantly in frequency over time with monthly intervals using Fisher's exact test. Another method is the sub-lineage corrected method for pVOI identification, which includes genome sequence assignments belonging to sublineages in the pVOI assessment. Finally, the sliding window method uses a windowed time period for the analysis to achieve a more fine-grained analysis of shorter time intervals (Methods). For each isolate, the provided Pangolin lineage[9] from the metadata is used as nomenclature. The

standard method calculates monthly frequencies for each lineage and detects significant increases in lineage frequency as described elsewhere[21]. The method utilizes Fisher's exact test to compare frequencies between consecutive months and corrects the false discovery rate (FDR) for multiple tests using the Benjamini-Hochberg procedure[22]. This identifies pVOIs as lineages that are both significantly on the rise and increase above a predominance threshold of 0.1 within the same time frame. Notably, lineages may also be falsely identified as pVOIs due to unrepresentative sampling, e.g., data biases towards certain areas, large clonal outbreaks, or population bottlenecks. Therefore, identified pVOIs should be evaluated carefully, potentially in combination with additional epidemiological data and experimental evidence showing that amino acid changes and altered positions observed in a pVOI lineage are likely to confer a selective advantage.

In silico assessment of antigenic lineage alterations

CoVerage predicts and visualizes the antigenic alteration of circulating lineages per country in a monthly map of the globe as well as a calculated antigenic alteration score per circulating Pango lineage. Current lineages with a score above 3.85 (defined based on lineages reported to have substantial antigenic alterations) are visualized in an antigenic alteration heatmap on the CoVerage platform. This displays lineages predicted to exhibit substantial antigenic alterations and their corresponding frequencies in different countries for the past month.

CoVerage lineage dynamics allow for timely identification of Omicron as pVOI

CoVerage provides lineage dynamics analyses for all countries with sufficient data available (set to more than 2000 sequences) to which one can navigate to via the interactive world map (Supplementary Fig. 1). This analysis is updated twice a week and suggests potential Variants of Interest for individual countries (**Fig. 2a**), as lineages rising in prevalence more rapidly than expected by chance (Methods), which may be due to a selective advantage and increased fitness to spread relative to other circulating viral lineages. To assess whether the lineage dynamics can be used to identify known VOCs from SARS-CoV-2 genomic surveillance data, we retrospectively analyzed data from the time of emergence of the Omicron lineage or Pango lineage BA.1. Using viral genome information provided until November 23rd of 2021, the day when the first Omicron sequences from South Africa identified under viral genomic surveillance[23] were submitted, the CoVerage lineage dynamic analysis suggested BA.1 as a pVOI, which then occurred with relative frequency of 0.27 among available South African isolates (**Fig. 2a**). The detection of BA.1 as a

pVOI was three days prior to the designation of Omicron as a VOC by the WHO[24]. The isolates used in the analysis were sampled between November 14th and November 16th, demonstrating how the methodology can rapidly identify novel lineages of concern when data are submitted to the GISAID database in a timely manner.

In addition to Omicron, the CoVerage lineage analysis of the South African data identified other lineages as pVOIs that were later designated VOCs by the WHO[25]; Beta (B.1.351) in September of 2020 and several Delta sublineages (AY.45, AY.38 and AY.32) throughout June and August of 2021. Additionally, two early lineages C.1 and B.1.1.54 were identified as pVOIs in June and July of 2020 in the standard and sublineage-corrected analyses. Both C.1 and B.1.1.54 rapidly increased in frequency among sequenced isolates from South Africa in the month of their designation but were later replaced with rising numbers of Beta. C.1, B.1.1.54, and a third lineage, B.1.156, comprised approximately 42% of the genomes sampled from the first wave in South Africa (from March to August of 2020), and by August 2020, the C.1 lineage was found to be the most geographically widespread lineage in South Africa[23]. Although never designated as a VOI or VOC by the WHO, CoVerage analysis identified the C.1 lineage as a pVOI a few months before August 2020.

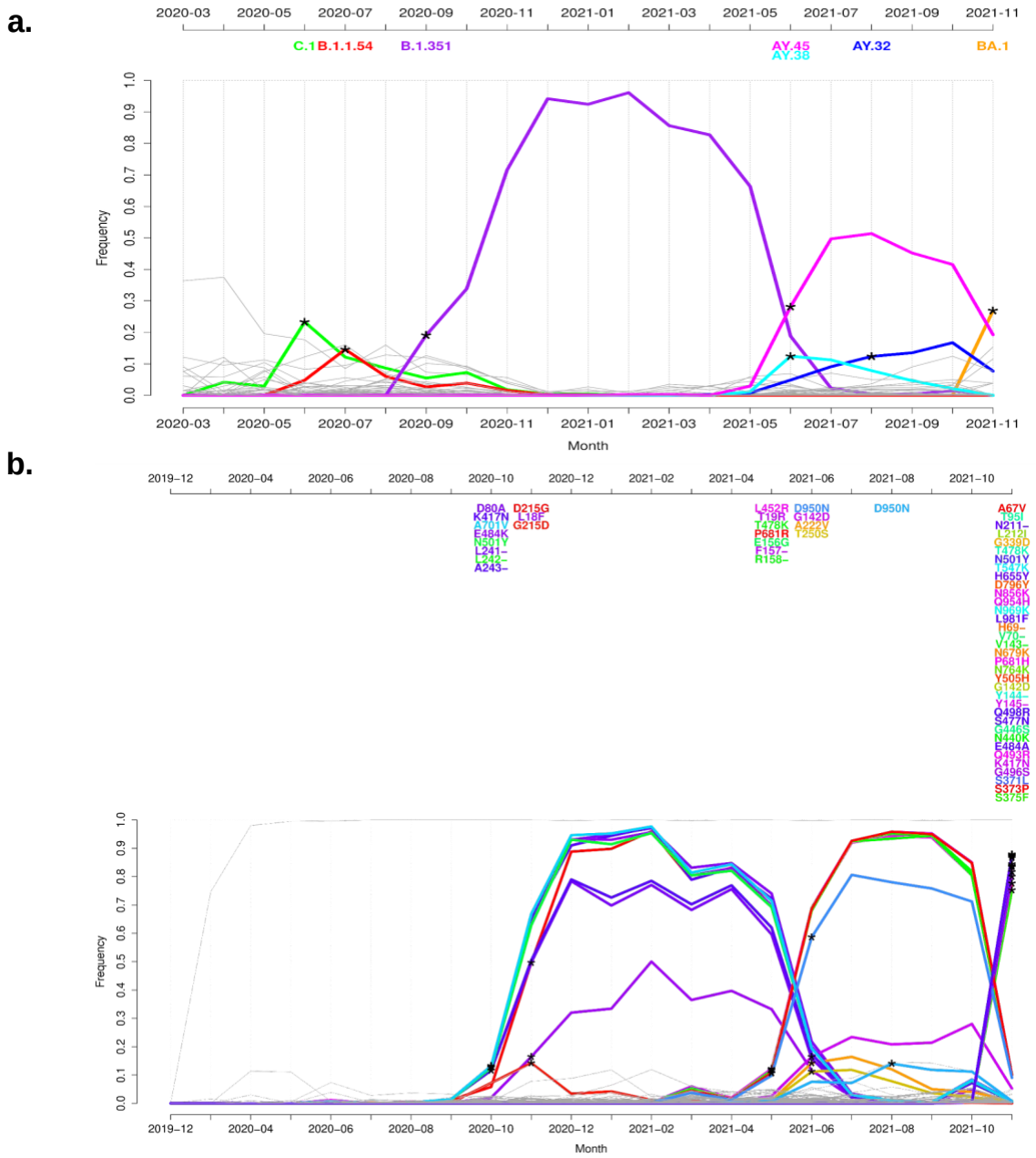

Fig. 2: a. Lineage dynamics plot for Pango lineages of SARS-CoV-2 genomes from South Africa submitted to GISAID by November 23rd, 2021. Colored lines represent identified pVOIs with their Pango lineage names given above. Asterisks indicate the month in which they were identified as being significantly on the rise and increasing in predominance above a predefined threshold of 0.1. All other lineages are shown in gray. **b. SD plot for South Africa on spike protein sequences available until the end of November 2021.** Similar to the lineage dynamics plot, the asterisks represent when a spike protein allele with certain amino acid changes significantly rises in frequency and the color of the curve corresponds to the associated amino

acid change. Amino acid changes are specified at the top of the plot, based on the time point when they were identified as significant. The asterisks in November 2021 indicate the detection of the emerging Omicron lineage (officially announced as a VOC on November 26th, 2021)[26], in May 2021 of Delta (Pango lineage B.1.617.2, detected with CoVerage first as a pVOI for India in March of 2021 and announced as a VOC on May 11th, 2021)[27], and Beta (B.1.351, announced as a VOC on November 29th in 2020) in October of 2020[28]. Asterisks in November 2020 indicate within lineage variation of Beta, in June and August of 2021 of sublineages AY.45, AY.38, and AY.32 of the Delta variant, respectively.

The relative frequencies of all pVOIs detected by CoVerage lineage dynamics analysis in countries worldwide with sufficient data are visualized weekly on the CoVerage website (**Fig. 3**). In November 2021, Omicron (BA.1) was first identified in this heatmap, initially primarily in South Africa (**Fig. 3a**), and then rose rapidly in frequency together with multiple sublineages such as BA.1.1, BA.1.15 and BA.1.17 to the most prevalent lineage worldwide by December 2021 (**Fig. 3b**).

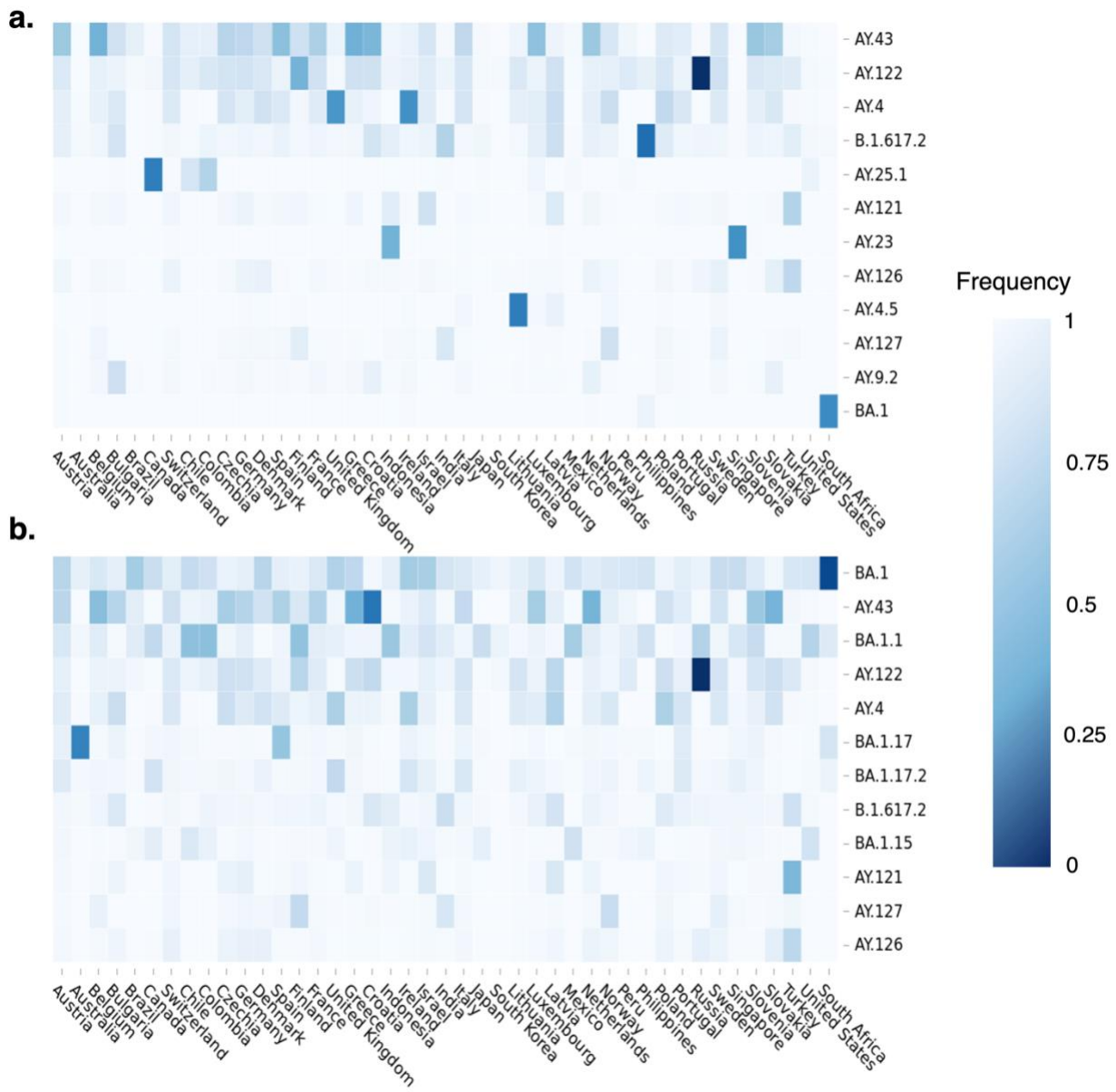

Fig. 3: Relative frequencies for SARS-CoV-2 pVOIs identified in individual countries and most abundant worldwide until a. November 2021 and b. December 2021. The color scale ranges from dark blue, which indicates a frequency of one, to white, indicating a frequency of zero. The top 50 countries with the highest lineage frequencies are shown.

Spike protein allele dynamics suggest emerging lineages and their associated amino acid changes

In addition, CoVerage provides an analysis of spike protein allele dynamics, which identifies “lineage alleles”, corresponding to branches in a viral spike protein genealogy and associated

amino acid changes increasing significantly in frequency over time, possibly because of them conferring a selective advantage to the respective lineages[21,29] (Methods).

To assess the value of this technique for identifying changes that were shown to provide a selective advantage experimentally, we analyzed sequences submitted in South Africa until the end of November 2021, to identify such amino acid changes for the Omicron lineage as well as for previously circulating lineages. This analysis detected all 35 substitutions representative of BA.1 in November 2021 (**Fig. 2b**), including changes shown to provide a selective advantage, as well as potential hitchhikers without phenotypic effect. More extensive sampling across the entire evolutionary process would be needed to determine which of the amino acid changes identified for Omicron were hitchhikers[30]. Among the changes known to alter the phenotype and provide a selective advantage are N501Y, which increases the binding affinity to the host angiotensin converting enzyme 2 (ACE2) receptor cells, E484A, which is a site relevant for immune escape, and P681H, which enhances spike cleavage[30,31].

Earlier, CoVerage also detected the Beta lineage and associated changes in October 2020, as well as Delta with its' amino acid changes in May 2021 (**Fig. 2b**). Key substitutions of the identified Beta lineage include the amino acid deletion at position 241-243, which is associated with reduced binding of certain nAbs[31], and K417N, that was also present in Alpha[31]. K417N, in combination with the E484K change in Beta, has demonstrated a reduced mAb neutralization, but does not display this capacity alone[10]. Key amino acid changes known to affect the phenotype of the Delta lineage identified in allele dynamics include L452R, which has previously been shown to be an important driver of adaptive evolution; T478K, which facilitates improved immune escape; and P681R, which enhances furin cleavage[30,31]. This demonstrates how the CoVerage allele dynamics analysis can be used to identify emerging VOCs and their distinctive amino acid alterations based on the ecological dynamics of the spike protein allele in the viral epidemic, even without their formal classification as novel lineages. This analysis facilitates a comprehensive and rapid characterization of relevant amino acid changes during the viral epidemic.

CoVerage antigenic alteration predictions

CoVerage provides an antigenic scoring analysis that identifies variants with predicted alterations in antigenicity based on changes at key amino acid positions on the SARS-CoV-2 spike protein. Lineages above the selected antigenic score threshold and their relative frequencies are shown as a heatmap (**Fig. 4**). To demonstrate the potential use of this methodology in identifying

antigenically altered lineages, we again focused on Omicron (Pango lineage BA.1) and its subsequent sublineages in November 2021. Omicron was assigned an antigenic alteration score of 3.72, a score over seven times greater than that of Delta (B.1.617.2; 0.5 for November 2021) and twice the antigenic score of Gamma (P.1; 2.16 for November 2021), which aligns with its' more pronounced antigenic change[32]. At this time, both Delta and Gamma were among the predominantly circulating lineages. Since December 2021 and continuing into January 2022, selected Omicron sublineages have already circulated globally at higher frequencies (**Fig. 4**). Some of these lineages, BA.2 and BA.2.10 both have even higher predicted antigenic scores than BA.1 in December (4.02 and 4.15, respectively).

On CoVerage, the rapid spread of Omicron and its sublineages can also be observed using the calculated antigenic drift scores of the currently circulating lineages per country, in the global antigenic change maps for October, November, and December 2021 (Methods). These scores reflect the circulating lineages, their antigenic alteration scores, and corresponding frequencies per country. For example, for South Africa, in November 2021 the country's antigenic score rose to 3.30 compared to 0.76 in the previous month (**Fig. 4d**). Throughout December 2021, many countries' antigenic scores increased, as Omicron and its sublineages rose in frequency, replacing previously circulating lineages (**Fig. 4e**).

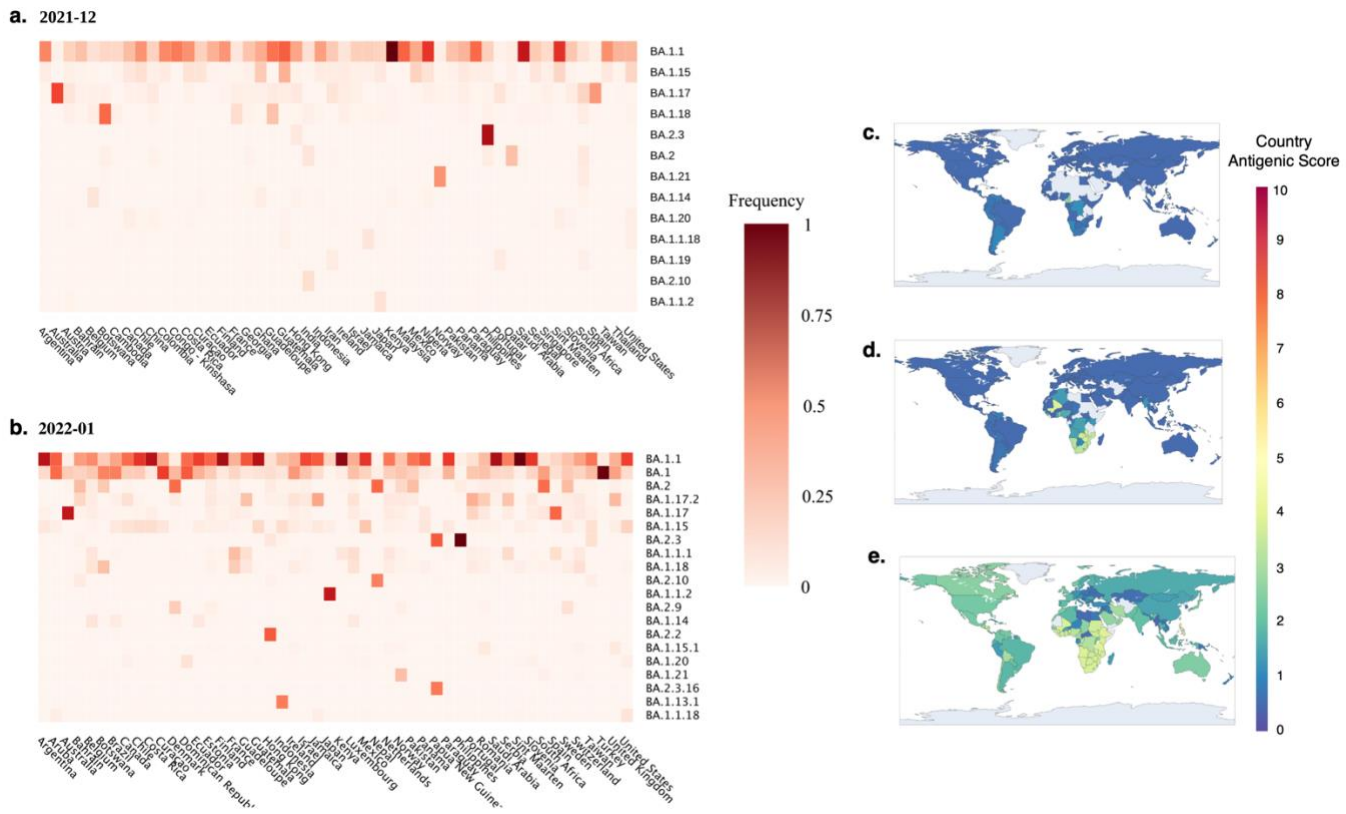

Fig. 4: Heatmap with lineages predicted to have substantial antigenic alteration relative to Wuhan-HU-1 and their frequencies per country as of a. December 2021 and b. January 2022. The 50 countries with the highest lineage frequencies are shown and the selected antigenically altered lineages are ordered from highest to lowest frequency. The dark red color indicates a frequency of 1, while the lighter color red represents a lower frequency closer to 0. **Global map depicting the assigned antigenic alteration score per country over the course of three months when Omicron became a predominant circulating VOC: c. October 2021, d. November 2021, and e. December 2021.** A higher score (warmer colors) means that antigenically altered lineages circulate with higher frequency in that country. Dark blue represents an antigenic score per country of 0, while red depicts an antigenic country antigenic score of 10.

Screening with CoVerage for current pVOIs with antigenic drift identifies relevant variants of interest weeks before their officially designation

To search for recent pVOIs and characterize them regarding their potential antigenic alterations and changes providing a selective advantage, we analyzed currently available data on March 31, 2023 with CoVerage (Supplementary Tables 3-5). In total, 48 lineages were identified as pVOIs

for January, 46 lineages in February, and 37 lineages in March. Most of these are sublineages of BQ.1 and XBB, denoted as BQ.1* and XBB* as per WHO specification[25]. In the final quarter of 2022 and in the beginning of 2023, BQ.1* was one of the major circulating lineages globally, with BQ.1 and BQ.1.1 together covering over 50 percent of submitted sequences in January 2023[33]. First identified in Nigeria[34], the WHO designated BQ.1 as a Variant of Interest on September 21, 2022. Using CoVerage, the BQ.1 lineage was identified as a pVOI in Nigeria in August 2022 - one month prior to its European Centre for Disease Prevention and Control (ECDC) designation as a Variant of Interest (**Fig. 5a**)[35]. Similarly, XBB.1.16 was identified by CoVerage as a pVOI in India in February 2023, two months prior to its VOI designation by the WHO on April 17, 2023[25] (**Fig. 5b**). Finally, XBB.1.5 was identified as a pVOI for the United States at the same time as its designation as a VOI by the WHO in January of 2023[25] (**Fig. 5c**).

In the *in silico* assessment of the antigenic drift of individual lineages, throughout the first quarter 2023, one of the highest-scoring circulating lineages was the CM.8 lineage. In March, CM.8 had an antigenic score of 6.98, nearly double the score of the original Omicron lineage (3.72). There are 35 changes in the spike protein of the CM.8 lineage relative to the Wuhan-Hu-1 strain, including ten changes (N440K, K444R, G446S, N450D, L452M, S477N, E484R, N501Y, P681H and D796Y) at antigenicity altering positions in the spike protein[36]. However, despite the high antigenic alteration score and numerous spike mutations, CM.8 had a low global cumulative prevalence during this time and was ultimately not identified as a pVOI[36]. We hypothesize that this could be due to competition with other variants, much like the case with Mu, which also showed signs of immune evasion, though Delta rose to predominance instead[37], or a loss of other phenotypically relevant properties, such as transmissibility. This also serves to demonstrate that substantial predicted antigenic alteration on its own is not predictive of a variant rising to predominance.

In March 2023, more BQ and XBB sublineages with antigenic alteration scores of more than 3.85 were identified (Supplementary Table 2). Both pVOIs BQ.1.1 and XBB.1.5 had an antigenic score of 5.69 and 4.72, respectively. In early 2023, BQ.1.1 initially increased to represent over 20% of global sequences but then was quickly overtaken by the XBB.1.5 lineage, which in March 2023 represented over 50% of global sequences[36]. In the antigenic alteration maps for individual lineages from January, February, and March 2023, we can see this transition from BQ.1.1 to XBB.1.5 as one of the major circulating lineages (**Fig. 5**). By February 6, 2023, XBB.1.5

represented 66.4% of sampled isolates in the USA and BQ.1.1 had a prevalence of 19.9%[38] (**Fig. 5**).

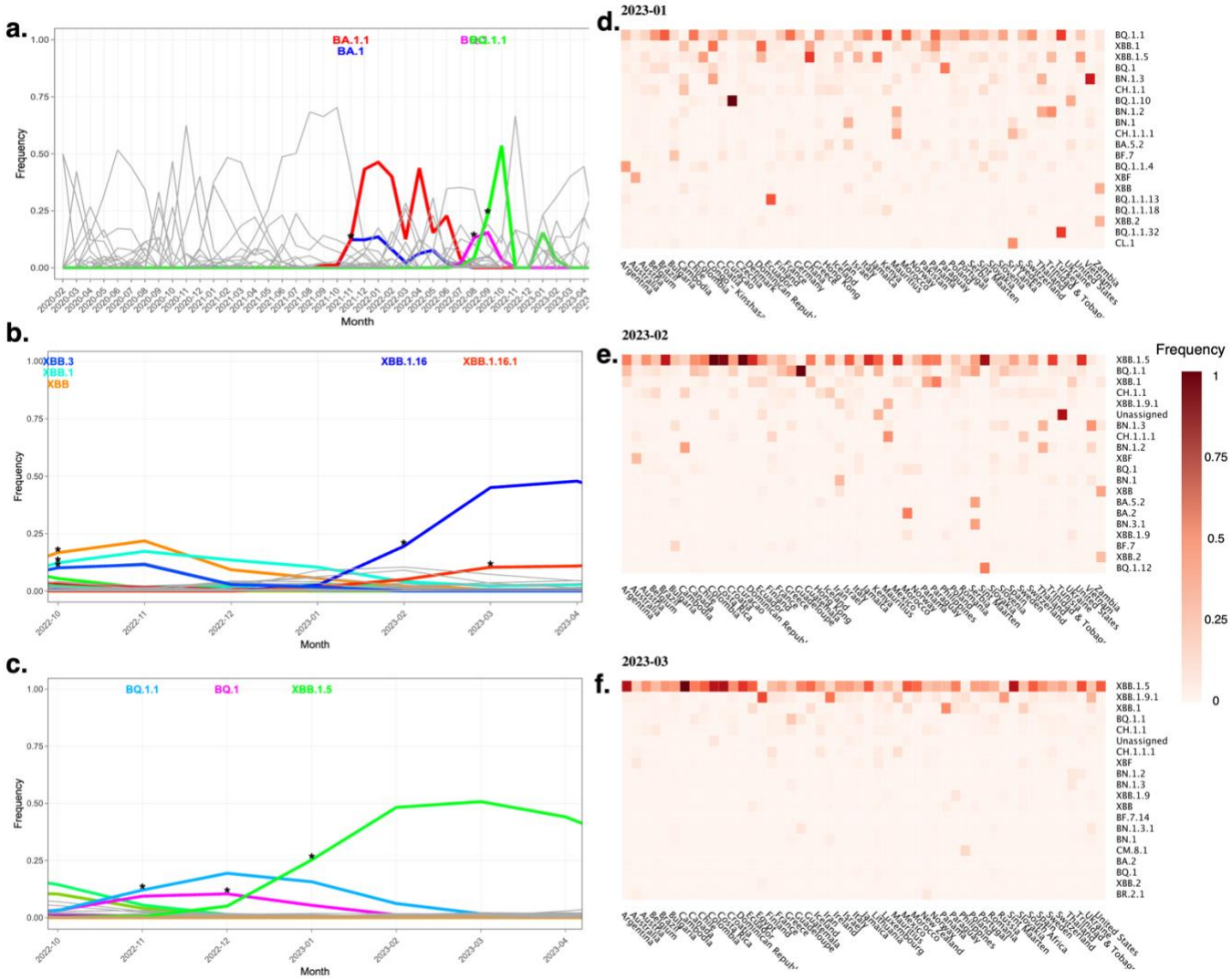

Fig. 5: Country-wise lineage dynamics plots for a. Nigeria, including results from April 2021 to April 2023, b. India, including results from October 2022 to April 2023, and c., the United States, including results from October 2022 to April 2023. a. The BQ.1 lineage is shown in pink, the BQ.1.1 lineage in red, and the BQ.1.1.18 lineage in purple. The asterisks represent a significant rise in frequency, which is seen in August 2022 (2022-08) for BQ.1. **b.** The XBB and XBB.1.16 lineages are shown in blue, XBB.1 in purple, and XBB.3 in green. The asterisk represents a significant rise in frequency for that lineage. **Heatmaps with antigenically altered lineages with an assigned antigenic alteration score above the predetermined threshold and their relative frequencies per country in a. January 2023 b. February 2023, and c. March**

2023. The lineages are ordered from highest to lowest frequency and the dark red color indicates a frequency of 1, while the lighter color red represents a lower frequency closer to 0.

The increased antigenic alteration scores of the Omicron BQ* and XBB* lineages are confirmed in their demonstrated capacity for immune evasion. Both BQ* and XBB* lineages have shown an increased immune evasion tendency greater than the other Omicron variants[33]. Key amino acid changes on the spike protein drive such improved immune escape, most notably the amino acid change F486P on the receptor binding domain of XBB.1.5[39], which also increases infectivity by improving the binding of the spike protein to host ACE2 receptor cells (Parums, 2023). Furthermore, the N460K mutation of BQ.1.1 has been associated with the loss of neutralizing activity of NTD-SD2 and class I mAbs, while the K444T and R346T changes in BQ.1.1 may also impair the potency of class III mAbs[33]. In May 2023, the WHO recommended an update of COVID-19 vaccine formulations to include the XBB.1.5 lineage, in line with the CoVerage results[40].

More recently, the JN.1 lineage, a BA.2.86 descendant, was designated a variant of interest by the WHO on December 18, 2023[41], two months after its demarcation as a pVOI in Portugal via the lineage dynamics analysis. Much like its predecessor, JN.1 carries over 30 mutations in the spike protein, one of the key amino acid changes being L455S[42]. This mutation was identified in December 2023 with the German allele dynamics analysis along with the L157S, N450D, L452W, and N481K amino acid changes. JN.1 subsequently showed extensive resistance to RBD class 1, 2, and 3 antibodies as well as monovalent XBB.1.5 vaccine sera[42,43] compared to BA.2.86 which is also reflected in its increased antigenic score of (6.18 and 5.84 respectively).

Discussion

Due to the ongoing and rapid genetic and antigenic evolution of circulating SARS-CoV-2 viruses, detecting emerging Variants of Concern that are on the rise to predominance as early as possible is crucial. This enables their timely in-depth assessment and initiating further actions, such as nonpharmaceutical interventions in the early stages of the pandemic and more recently, updating the vaccine composition to ensure continued effectiveness[40]. Here, we describe CoVerage, a genomic surveillance platform for continuously monitoring globally provided SARS-CoV-2 genomics data, to identify and characterize potential Variants of Interest from the circulating SARS-CoV-2 lineage diversity via *in silico* techniques. CoVerage implements three types of innovative methods for this purpose: a method for the *de novo* detection of potential Variants of

Interest that may spread more efficiently than others, a method to identify amino acid changes in the major surface spike protein that may confer a selective advantage, and a method for scoring the degree of antigenic alteration of individual sequences, lineages and the circulating viral diversity per country. To ensure maximal relevance for viral surveillance, CoVerage provides up-to-date predictions of current pVOIs and their antigenic alterations for all countries around the globe with sufficient data available twice a week. Using the emergence of Omicron, the most recent variant of concern, and its subsequently established sublineages as examples, we demonstrate how CoVerage can be applied to rapidly detect these as pVOIs, to assess their antigenic differences, and track their ongoing, global spread.

CoVerage provides functionalities that extend beyond the surveillance of circulating lineages and their characteristics as available before[9,13,15]. As we show for the example of Omicron and relevant sublineages, the *de novo* suggestion of potential Variants of Interest allows for a timely identification of novel, emerging VOIs with CoVerage. For example, within this framework, Omicron was identified as a pVOI shortly prior to its WHO designation as a Variant of Concern in November of 2021, and XBB 1.5, which was recommended for a vaccine update in May 2023[40], was identified as early as January 2023. In terms of data, CoVerage draws on the international GISAID data and combines it with other resources where available, such as the genome and metadata published on GitHub/Zenodo by the German Robert Koch Institute. By incorporating online resources where genomic surveillance data is available more quickly than on GISAID, CoVerage aims to further decrease the time to detecting new, relevant variants[44].

Currently, the pVOIs suggested by CoVerage for any point in time worldwide cover a dozen lineages and their sublineages, that largely represent subsequently described VOIs and VOCs. Notably, detection quality depends on the extent and quality of ongoing viral genomic surveillance programs for individual countries, and may also be affected by population genetic effects[45], such as population bottlenecks, when case numbers are low, or travel restrictions between countries are in place, such as in early phases of the pandemic. Focusing on pVOIs identified simultaneously in many countries, including those with in-depth genomic surveillance programs using the country-wise lineage dynamics plots provided by CoVerage, can help to obtain a greater level of confidence for VOI identification and to mitigate the influence of these varying effects.

Different from other resources (e.g. CovidCG[13]), which monitor amino acid changes from sequences, in selecting lineages or visualize positions in the spike protein to provide a general

overview, the CoVerage amino acid dynamics analysis identifies alleles of the major surface protein of SARS-CoV-2 carrying amino acid substitutions that may provide a selective advantage via a spike protein genealogy phylogenetic tree. This is a helpful technique for identifying candidate amino acid changes for particular lineages that confer phenotypic alterations, such as in antigenicity, transmissibility or viral replication[21,29]. For example, N501Y was detected in September of 2020 for Alpha and then again in November of 2021 for Omicron in South Africa, and is associated with improved immune escape[31]. The method allows us to distinguish between shared substitutions that have only been acquired once, such as D614G that occurs in all VOCs and evolved at the beginning of the pandemic, and substitutions that have been acquired independently multiple times (convergent evolution), and to monitor the evolutionary trajectories of individual spike protein "alleles" over time, which has been adapted from an approach to identify suitable strains of seasonal influenza A viruses for vaccine strain updates[46].

Finally, the *in silico* assessment of antigenic lineage alterations analysis scores lineages based on amino acid changes occurring throughout the SARS-CoV-2 spike protein and their potential impact on the virus' capacity for immune escape. Using a selected set of amino acid sites that were shown to alter viral antigenic phenotype, we scored variants occurring in global GISAID sequences based on amino acid changes occurring at these sites. To test this methodology, we tracked the initial spread of Omicron and its early sublineages throughout the last few months of 2021 and early 2022. As of November 2021, we note an increase in South Africa's country antigenic score, and upon closer inspection with the antigenic heatmap found numerous circulating VOIs, including Omicron and its initial sub lineages. These variants included both BA.1 and BA.2 sublineages in December of 2021 and January of 2022. The methodology was then used to identify antigenically altered lineages in the first quarter of 2023, using GISAID metadata from April 2023. The majority of selected antigenically altered lineages were sublineages of BQ.1 and XBB, both of which have been shown to have improved immune escape due to key mutations throughout the RBD and NTD[33]. Different from other methods predicting viral antigenicity[47], the technique provided by CoVerage scores variants' potential antigenicity based on only genomic metadata and can be used to compare lineages against one another or track the virus' change in antigenicity over time. The greater immune escape of the lineages BQ* and XBB* are reflected in their antigenic drift scores and together with their recent, rapid increase in frequencies per the lineage dynamics analysis and pVOI designation from late 2022 onwards, preceded the suggestion of XBB.1.5 as vaccine strain[40] in May 2023 by the WHO. Taken together, this shows that the combination of these analyses provided by CoVerage allows for the efficient identification

of antigenically altered pVOI's. More recently, in November 2023, pVOI HK.3 (XBB.1.9.2.5.1.1.3), together with other sublineages of EG.5, is predicted as the most abundant pVOI worldwide, which was first identified on CoVerage as a pVOI in China as of August 2023, and predicted to be more antigenically altered than XBB.1.5, in agreement with first scientific reports (with an antigenic alteration score of 4.95 versus 4.8 respectively)[48].

Our method allows us to monitor and compare lineages and their potential antigenicity as measured by their score. The global map can be used to track changes in each country as different lineages rise in frequency while the heatmap allows for deeper investigation. We note with Omicron a substantial increase in antigenic score from previously circulating lineages such as Delta and Gamma. Though slightly delayed in identifying variants, when this analysis is used in combination with the previous analyses we can get a comprehensive overview of antigenically altered lineages, together with potential lineages of interest rapidly rising in frequency.

Taken together, CoVerage is a unique web-based resource to identify potential Variants of Interest of SARS-CoV-2 in a timely manner, along with suggesting their degree of antigenic alteration, and alleles of the major surface protein with specific amino acid changes that may provide a selective advantage. Its functionalities complement other pathogen surveillance web-based resources such as NextStrain[14], which established a viral lineage nomenclature based on phylogenetic principles for SARS-CoV-2 and infers phylogenies on subsampled datasets to visualize the genomic evolution of SARS-CoV-2, while displaying the prevalence of different viral clades both globally and per continent; CoVariants[15] that provides an overview of variants and mutations of interest, which are also linked to academic literature about their potential antigenic impact; CovidCG[13], a web-based tool that can quickly track mutations, lineages, and clades present in an area during a specific period without subsampling; outbreak.info[17], which aggregates data to track daily developments in SARS-CoV-2 variants, provides an in-depth review of each variant including mutation sites and current prevalence, tracks cases and deaths, as well as integrates SARS-CoV-2 literature into a searchable library format; CovRadar[18], a platform-specific to Germany for the molecular surveillance of the SARS-CoV-2 spike protein; and the COVID-19 Viral Genome Analysis Pipeline[16], which identifies mutations on the SARS-CoV-2 spike protein that are increasing in frequency across different geographic regions. Further resources[13,15-17] provide information about lineages and amino acid changes directly from sequence data, and include the prevalence of variants, their geographical distributions, as well as comparisons of changes between lineages. It should also be noted that many of these

resources have additional functions that may not be outlined in entirety here. However, CoVerage is unique in that it predicts potential Variants of Interest, which have also been identified as antigenically altered, oftentimes prior to the WHO designation. These lineages can be further investigated using alternative web-based resources that Coverage links to directly for individual lineages in multiple places, yet the key to CoVerage is identifying which variants are of importance for further monitoring within the current Omicron "variant swarm"[49]. Each of the different types of analyses provided on CoVerage offers its own benefits when used individually, but when used in combination with each other, is even more powerful, such as the ability to identify pVOIs that contain potentially antigenically altering changes on the rise to predominance, which therefore may be relevant to consider for vaccine strain updates.

CoVerage provides a continuously updated resource for the *in silico* detection and analysis of pVOIs from SARS-CoV-2 genomic surveillance data, to support researchers and public health officials in assessing and interpreting such data. The framework can be extended to provide relevant surveillance for other rapidly evolving viruses, such as seasonal influenza viruses, and integrate further data types, such as viral genome recovery from metagenomic sequencing of wastewater samples[50].

Methods

Lineage dynamics by region

In the sublineage corrected analysis, the lineage frequencies are corrected by including isolates belonging to sublineages in the count, and subsequently processed as before. For example, three sublineages BA.1.2, BA.1.3 and BA.1.4 will be summarized into BA.1 and then analyzed with Fisher's exact test. This approach leads to a better resolution of the lineage branches likely associated with more rapid spread and lineages with a selective advantage split into several sublineages can be better detected.

Lastly, a sliding window approach was implemented to achieve a more sensitive detection than the monthly analysis. For this analysis, sequences are sorted by date and frequencies in the w sequences in the current window are compared to the w previous ones using Fisher's exact test. The window is moved over the data using a step size s. Significance estimates are corrected for multiple testing using the Benjamini-Yekutieli procedure[51]. For the analysis on countries, $w =$

1000 and $s = 100$ are used and for the analysis on more granular German state level $w = 200$ and $s = 10$ is chosen. Additionally, we require a significant (FDR < 0.05) increase for a certain window and a frequency threshold of 0.1 to report results. The date of the last sequence in the current window is reported as the date the significant increase is detected.

Spike protein allele dynamics

To identify amino acid changes in the spike protein that may provide a lineage with a selective advantage, the sweep dynamic (SD) plot method[21,29] is used on spike protein sequences extracted from viral genome sequences data downloaded from GISAID and Zenodo. To execute this methodology on the large SARS-CoV-2 sequence collection, 500 sequences per month per country are subsampled randomly and downloaded with GISAIDR[52], and subsequently, identical sequences per time period are clustered using CD-HIT[53]. For German state-wise analysis, we randomly downloaded 2000 sequences per month for the entire country and divided them by states. Next, a multiple sequence alignment is generated using MAFFT[54], with the spike sequence of Wuhan/IPBCAMS-WH-01/2019 as a reference. Phylogenetic trees are inferred for each country using fasttree[55], and the Sankoff algorithm is applied for ancestral character state reconstruction, using the same cost for all amino acid changes and indels. The calculation of relative frequencies of specific spike protein alleles and statistical evaluation are then performed as described in[21,29], with a frequency threshold of 0.1 for finding spike protein alleles and their amino acid changes that may provide a selective advantage.

***In silico* assessment of antigenic lineage alterations**

The antigenic alteration score for a lineage a_l is calculated using a method we developed to score each SARS-CoV-2 spike protein sequence relative to the Wuhan-Hu-1 strain (attached manuscript in preparation). In addition to the individual lineage score, a country-wide antigenic alteration score a_c is calculated by summing over all antigenic lineage scores, weighted by their frequencies for a particular period in time:

$$a_c = \sum(a_l \times \frac{n_{ilc}}{n_{ic}})$$

Here, n_{ilc} is the number of isolates of each Pango lineage for country *c* and n_{ic} is the total number of isolates per country.

References

1. Wu, F., Zhao, S., Yu, B., Chen, Y. M., Wang, W., Song, Z. G., Hu, Y., Tao, Z. W., Tian, J. H., Pei, Y. Y., Yuan, M. L., Zhang, Y. L., Dai, F. H., Liu, Y., Wang, Q. M., Zheng, J. J., Xu, L., Holmes, E. C., & Zhang, Y. Z. (2020). A new coronavirus associated with human respiratory disease in China. *Nature*, *579*(7798), 265–269. https://doi.org/10.1038/s41586-020-2008-3
2. Dong, E., Du, H., & Gardner, L. (2020). An interactive web-based dashboard to track COVID-19 in real time. In *The Lancet Infectious Diseases* (Vol. 20, Issue 5, pp. 533–534). Lancet Publishing Group. https://doi.org/10.1016/S1473-3099(20)30120-1
3. Ortiz-Prado, E., Simbaña-Rivera, K., Gómez- Barreno, L., Rubio-Neira, M., Guaman, L. P., Kyriakidis, N. C., Muslin, C., Jaramillo, A. M. G., Barba-Ostria, C., Cevallos-Robalino, D., Sanches-SanMiguel, H., Unigarro, L., Zalakeviciute, R., Gadian, N., & López-Cortés, A. (2020). Clinical, molecular, and epidemiological characterization of the SARS-CoV-2 virus and the Coronavirus Disease 2019 (COVID-19), a comprehensive literature review. In *Diagnostic Microbiology and Infectious Disease* (Vol. 98, Issue 1). Elsevier Inc. https://doi.org/10.1016/j.diagmicrobio.2020.115094
4. Yao, H., Song, Y., Chen, Y., Wu, N., Xu, J., Sun, C., Zhang, J., Weng, T., Zhang, Z., Wu, Z., Cheng, L., Shi, D., Lu, X., Lei, J., Crispin, M., Shi, Y., Li, L., & Li, S. (2020). Molecular Architecture of the SARS-CoV-2 Virus. *Cell*, *183*(3), 730-738.e13. https://doi.org/10.1016/j.cell.2020.09.018
5. Davies, N. G., Jarvis, C. I., van Zandvoort, K., Clifford, S., Sun, F. Y., Funk, S., Medley, G., Jafari, Y., Meakin, S. R., Lowe, R., Quaife, M., Waterlow, N. R., Eggo, R. M., Lei, J., Koltai, M., Krauer, F., Tully, D. C., Munday, J. D., Showering, A., … Keogh, R. H. (2021). Increased mortality in community-tested cases of SARS-CoV-2 lineage B.1.1.7. *Nature*, *593*(7858), 270–274. https://doi.org/10.1038/s41586-021-03426-1
6. Faria, Nuno R., Thomas A. Mellan, Charles Whittaker, Ingra M. Claro, Darlan da S. Candido, Swapnil Mishra, Myuki A. E. Crispim, et al. 2021. “Genomics and Epidemiology of the P.1 SARS-CoV-2 Lineage in Manaus, Brazil.” *Science* 372 (6544): 815–21.
7. Garcia-Beltran, Wilfredo F., Evan C. Lam, Kerri St Denis, Adam D. Nitido, Zeidy H. Garcia, Blake M. Hauser, Jared Feldman, et al. 2021. “Multiple SARS-CoV-2 Variants Escape Neutralization by Vaccine-Induced Humoral Immunity.” *Cell* 184 (9): 2372–83.e9.
8. Volz, Erik, Swapnil Mishra, Meera Chand, Jeffrey C. Barrett, Robert Johnson, Lily Geidelberg, Wes R. Hinsley, et al. 2021. “Assessing Transmissibility of SARS-CoV-2 Lineage B.1.1.7 in England.” *Nature* 593 (7858): 266–69.
9. Rambaut, A., Holmes, E. C., O’Toole, Á., Hill, V., McCrone, J. T., Ruis, C., du Plessis, L., & Pybus, O. G. (2020). A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. *Nature Microbiology*, *5*(11), 1403–1407. https://doi.org/10.1038/s41564-020-0770-5
10. Cox, M., Peacock, T. P., Harvey, W. T., Hughes, J., Wright, D. W., Willett, B. J., Thomson, E., Gupta, R. K., Peacock, S. J., Robertson, D. L., & Carabelli, A. M. (2022). SARS-CoV-2 variant evasion of monoclonal antibodies based on in vitro studies. *Nature Reviews Microbiology 2022*, 1–13. https://doi.org/10.1038/s41579-022-00809-7
11. Rössler, A., Knabl, L., von Laer, D., & Kimpel, J. (2022). Neutralization Profile after Recovery from SARS-CoV-2 Omicron Infection. *New England Journal of Medicine*, *386*(18), 1764–1766.https://doi.org/10.1056/NEJMC2201607/SUPPL_FILE/NEJMC2201607_DISCLOSURES.PDF
12. Shu, Yuelong, and John McCauley. 2017. “GISAID: Global Initiative on Sharing All Influenza Data – from Vision to Reality.” *Euro Surveillance: Bulletin Europeen Sur Les Maladies Transmissibles = European Communicable Disease Bulletin* 22 (13). https://doi.org/10.2807/1560-7917.es.2017.22.13.30494.

13. Chen, Albert Tian, Kevin Altschuler, Shing Hei Zhan, Yujia Alina Chan, and Benjamin E. Deverman. 2021. “COVID-19 CG Enables SARS-CoV-2 Mutation and Lineage Tracking by Locations and Dates of Interest.” *eLife* 10 (February). https://doi.org/10.7554/eLife.63409.
14. Hadfield, James, Colin Megill, Sidney M. Bell, John Huddleston, Barney Potter, Charlton Callender, Pavel Sagulenko, Trevor Bedford, and Richard A. Neher. 2018. “Nextstrain: Real-Time Tracking of Pathogen Evolution.” *Bioinformatics* 34 (23): 4121–23.
15. Hodcroft, Emma B. 2021. “CoVariants.” CoVariants: SARS-CoV-2 Mutations and Variants of Interest. 2021. https://covariants.org/.
16. Korber, Bette, Will M. Fischer, Sandrasegaram Gnanakaran, Hyejin Yoon, James Theiler, Werner Abfalterer, Nick Hengartner, et al. 2020. “Tracking Changes in SARS-CoV-2 Spike: Evidence That D614G Increases Infectivity of the COVID-19 Virus.” *Cell* 182 (4): 812–27.e19.
17. Mullen, Julia L., Ginger Tsueng, Alaa Abdel Latif, Manar Alkuzweny, Marco Cano, Emily Haag, Jerry Zhou, et al. 2020. “Outbreak.info.” Outbreak.info. 2020. https://outbreak.info/.
18. Wittig, A., Miranda, F., Hölzer, M., Altenburg, T., Bartoszewicz, J. M., Beyvers, S., Dieckmann, M. A., Genske, U., Giese, S. H., Nowicka, M., Richard, H., Schiebenhoefer, H., Schmachtenberg, A. J., Sieben, P., Tang, M., Tembrockhaus, J., Renard, B. Y., & Fuchs, S. (2022). CovRadar: Continuously tracking and filtering SARS-CoV-2 mutations for genomic surveillance. *Bioinformatics*, *38*(17), 4223–4225. https://doi.org/10.1093/bioinformatics/btac411
19. Khare, S., Gurry, C., Freitas, L., Schultz, MRapid and parallel adaptive mutations in spike S1 drive clade success in SARS-CoV-2. B., Bach, G., Diallo, A., Akite, N., Ho, J., Lee, R. T., Yeo, W., Curation Team, G. C., & Maurer-Stroh, S. (2021). GISAID's Role in Pandemic Response. *China CDC weekly*, *3*(49), 1049–1051. https://doi.org/10.46234/ccdcw2021.255
20. WHO COVID-19 Dashboard. Geneva: World Health Organization, 2020. Available online: https://covid19.who.int/ (last cited: [22-11-2023])
21. Klingen, T. R., Reimering, S., Loers, J., Mooren, K., Klawonn, F., Krey, T., Gabriel, G., & Mchardy, A. C. (2018). Sweep Dynamics (SD) plots: Computational identification of selective sweeps to monitor the adaptation of influenza A viruses. *ScIentIfIc RepoRTS* |, *8*, 373. https://doi.org/10.1038/s41598-017-18791-z
22. Benjamini, Y., & Hochberg, Y. (1995). Controlling the False Discovery Rate: a Practical and Powerful Approach to Multiple Testing. In *J. R. Statist. Soc. B* (Vol. 57, Issue I).
23. Viana, R., Moyo, S., Amoako, D. G., Tegally, H., Scheepers, C., Althaus, C. L., Anyaneji, U. J., Bester, P. A., Boni, M. F., Chand, M., Choga, W. T., Colquhoun, R., Davids, M., Deforche, K., Doolabh, D., du Plessis, L., Engelbrecht, S., Everatt, J., Giandhari, J., … de Oliveira, T. (2022). Rapid epidemic expansion of the SARS-CoV-2 Omicron variant in southern Africa. *Nature*, *603*(7902), 679–686. https://doi.org/10.1038/s41586-022-04411-y
24. World Health Organization. (2021, November 28). *Update on Omicron.* Retrieved from https://www.who.int/news/item/28-11-2021-update-on-omicron
25. World Health Organization. (2022, October 27). *Tracking SARS-CoV-2 variants.* Retrieved from https://www.who.int/activities/tracking-SARS-CoV-2-variants
26. He, X., Hong, W., Pan, X., Lu, G., & Wei, X. (2021). SARS-CoV-2 Omicron variant: Characteristics and prevention. In *MedComm* (Vol. 2, Issue 4, pp. 838–845). Blackwell Publishing Inc. https://doi.org/10.1002/mco2.110
27. Dhawan, M., Sharma, A., Priyanka, Thakur, N., Rajkhowa, T. K., & Choudhary, O. P. (2022). Delta variant (B.1.617.2) of SARS-CoV-2: Mutations, impact, challenges and possible solutions. In *Human Vaccines and Immunotherapeutics* (Vol. 18, Issue 5). Taylor and Francis Ltd. https://doi.org/10.1080/21645515.2022.2068883
28. Center for Disease Control and Prevention. (2023). *SARS-CoV-2 Variant Classifications and Definitions*. National Center for Immunization and Respiratory Diseases (NCIRD), Division of Viral Diseases. https://www.cdc.gov/coronavirus/2019-ncov/variants/variant-classifications.html
29. Steinbrück, L., & McHardy, A. C. (2011). Allele dynamics plots for the study of evolutionary dynamics in viral populations. *Nucleic Acids Research*, *39*(1). https://doi.org/10.1093/nar/gkq909

30. Kistler, K. E., Huddleston, J., & Bedford, T. (2022). Rapid and parallel adaptive mutations in spike S1 drive clade success in SARS-CoV-2. *Cell Host and Microbe*, *30*(4), 545-555.e4. https://doi.org/10.1016/j.chom.2022.03.018
31. Mittal, A., Khattri, A., & Verma, V. (2022). Structural and antigenic variations in the spike protein of emerging SARS-CoV-2 variants. *PLOS Pathogens*, *18*(2), e1010260. https://doi.org/10.1371/JOURNAL.PPAT.1010260
32. Telenti, A., Hodcroft, E. B., & Robertson, D. L. (2022). The Evolution and Biology of SARS-CoV-2 Variants. *Cold Spring Harbor Perspectives in Medicine*, *12*(5). https://doi.org/10.1101/cshperspect.a041390
33. Ao, D., He, X., Hong, W., & Wei, X. (2023). The rapid rise of SARS-CoV-2 Omicron subvariants with immune evasion properties: XBB.1.5 and BQ.1.1 subvariants. *MedComm*, *4*(2). https://doi.org/10.1002/mco2.239
34. Akash, S., Islam, Md. R., & Dhama, K. (2023). Emergence BQ.1 and BQ.1.1 as newly identified omicron subvariants: current scenario and future outlook – an update. *Annals of Medicine & Surgery*, *85*(4), 1329–1330. https://doi.org/10.1097/MS9.0000000000000469
35. European Centre for Disease Prevention and Control. (October 21, 2022). *Spread of the SARS-CoV-2 sublineage in the EU/EEA*. https://www.ecdc.europa.eu/en/publications-data/spread-sars-cov-2-omicron-variant-sub-lineage-bq1-eueea#:~:text=Executive%20summary-,BQ.,1%2FBQ.l
36. Outbreak.info genomic reports: scalable and dynamic surveillance of SARS-CoV-2 variants and mutations. Karthik Gangavarapu, Alaa Abdel Latif, Julia L. Mullen, Manar Alkuzweny, Emory Hufbauer, Ginger Tsueng, Emily Haag, Mark Zeller, Christine M. Aceves, Karina Zaiets, Marco Cano, Xinghua Zhou, Zhongchao Qian, Rachel Sattler, Nathaniel L. Matteson, Joshua I. Levy, Raphael T. C. Lee, Lucas Freitas, Sebastian Maurer-Stroh, GISAID Core and Curation Team, Marc A. Suchard, Chunlei Wu, Andrew I. Su, Kristian G. Andersen & Laura D. Hughes. Nature Methods (2023). doi: 10.1038/s41592-023-01769-3
37. Petrone, M. E., Lucas, C., Menasche, B., Breban, M. I., Yildirim, I., Campbell, M., Omer, S. B., Ko, A. I., Grubaugh, N. D., Iwasak, A., Wilen, C. B., Vogels, C. B. F., & Fauver, J. R., (2022). Insights into the limited global spread of the immune evasive SARS-CoV-2 variant Mu. *medRxiv : the preprint server for health sciences*, 2022.03.28.22273077. https://doi.org/10.1101/2022.03.28.22273077
38. Zhu, A., Wei, P., Man, M., Liu, X., Ji, T., Chen, J., Chen, C., Huo, J., Wang, Y., & Zhao, J. (2023). Antigenic characterization of SARS-CoV-2 Omicron subvariants XBB.1.5, BQ.1, BQ.1.1, BF.7 and BA.2.75.2. *Signal Transduction and Targeted Therapy*, *8*(1). https://doi.org/10.1038/s41392-023-01391-x
39. Greaney, A. J., Starr, T. N., Barnes, C. O., Weisblum, Y., Schmidt, F., Caskey, M., Gaebler, C., Cho, A., Agudelo, M., Finkin, S., Wang, Z., Poston, D., Muecksch, F., Hatziioannou, T., Bieniasz, P. D., Robbiani, D. F., Nussenzweig, M. C., Bjorkman, P. J., & Bloom, J. D. (2021). Mapping mutations to the SARS-CoV-2 RBD that escape binding by different classes of antibodies. *Nature Communications*, *12*(1). https://doi.org/10.1038/s41467-021-24435-8
40. World Health Organization. (2023, May 18). *Statement on the antigenic composition of COVID-19 vaccines.* Retrieved from https://www.who.int/news/item/18-05-2023-statement-on-the-antigen-composition-of-covid-19-vaccines
41. Jeworowski, L. M., Mühlemann, B., Walper, F., Schmidt, M. L., Jansen, J., Krumbholz, A., Simon-Lorière, E., Jones, T. C., Corman, V. M., & Drosten, C. (2024). Humoral immune escape by current SARS-CoV-2 variants BA.2.86 and JN.1, December 2023. Euro Surveillance : Bulletin Europeen Sur Les Maladies Transmissibles = European Communicable Disease Bulletin, 29(2). https://doi.org/10.2807/1560-7917.ES.2024.29.2.2300740
42. Kaku, Y., Okumura, K., Padilla-Blanco, M., Kosugi, Y., Uriu, K., Hinay, A. A., Chen, L., Plianchaisuk, A., Kobiyama, K., & Ishii, K. J. (2024). Virological characteristics of the SARS-CoV-2 JN.1 variant. The Lancet Infectious Diseases, 24(2). https://doi.org/https://doi.org/10.1016/S1473-3099(23)00813-7

43. Yang, S., Yu, Y., Xu, Y., Jian, F., Song, W., Yisimayi, A., Wang, P., Wang, J., Liu, J., Yu, L., Niu, X., Wang, J., Wang, Y., Shao, F., Jin, R., Wang, Y., & Yunlong. (2023). Fast evolution of SARS-CoV-2 BA.2·86 to JN.1 under heavy immune pressure. The Lancet Infectious Diseases, 24(2). https://doi.org/10.1016/s1473-3099(23)00744-2
44. Kalia, K., Saberwal, G., & Sharma, G. (2021). The lag in SARS-CoV-2 genome submissions to GISAID. In *Nature Biotechnology* (Vol. 39, Issue 9, pp. 1058–1060). Nature Research. https://doi.org/10.1038/s41587-021-01040-0
45. Martin, M. A., VanInsberghe, D., & Koelle, K. (2021). Insights from SARS-CoV-2 sequences. *Science*, *371*(6528), 466–467.
46. Steinbrück, L., Klingen, T. R., & McHardy, A. C. (2014). Computational Prediction of Vaccine Strains for Human Influenza A (H3N2) Viruses. *Journal of Virology*, *88*(20), 12123–12132. https://doi.org/10.1128/jvi.01861-14
47. Pucci, F., & Rooman, M. (2021). Prediction and evolution of the molecular fitness of sars-cov-2 variants: Introducing spikepro. *Viruses*, *13*(5). https://doi.org/10.3390/v13050935
48. Kosugi, Y., Plianchaisuk, A., Putri, O., Uriu, K., Kaku, Y., Hinay Jr, A. A., Chen, L., Kuramochi, J., Sadamasu, K., Asakura, H., Nagashima, M., Ito, J., & Sato, K. (2023). Virological characteristics of the SARS-CoV-2 Omicron HK.3 variant 1 harboring the “FLip” substitution 2 3. BioRxiv. https://doi.org/10.1101/2023.11.14.566985
49. Cao, Y., Yisimayi, A., Jian, F., Song, W., Xiao, T., Wang, L., Du, S., Wang, J., Li, Q., Chen, X., Yu, Y., Wang, P., Zhang, Z., Liu, P., An, R., Hao, X., Wang, Y., Wang, J., Feng, R., … Xie, X. S. (2022). BA.2.12.1, BA.4 and BA.5 escape antibodies elicited by Omicron infection. *Nature*, *608*(7923), 593–602. https://doi.org/10.1038/s41586-022-04980-y
50. Fritz, A., Bremges, A., Deng, Z. L., Lesker, T. R., Götting, J., Ganzenmueller, T., Sczyrba, A., Dilthey, A., Klawonn, F., & McHardy, A. C. (2021). Haploflow: strain-resolved de novo assembly of viral genomes. *Genome Biology*, *22*(1). https://doi.org/10.1186/s13059-021-02426-8
51. Benjamini, Yoav, and Daniel Yekutieli. 2001. “The Control of the False Discovery Rate in Multiple Testing under Dependency.” *The Annals of Statistics* 29 (4): 1165–88.
52. Wytamma Wirth, & Sebastian Duchene. (2022). GISAIDR: Programmatically interact with the GISAID databases. Zenodo. https://doi.org/10.5281/zenodo.6474693
53. Limin Fu, Beifang Niu, Zhengwei Zhu, Sitao Wu and Weizhong Li, CD-HIT: accelerated for clustering the next generation sequencing data. Bioinformatics, (2012), 28 (23): 3150-3152. doi: 10.1093/bioinformatics/bts565. PDF, pubmed.
54. Katoh, K., Misawa, K., Kuma, K.-I., & Miyata, T. (2002). *MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform*.
55. Price, M.N., Dehal, P.S., and Arkin, A.P. (2010) FastTree 2 -- Approximately Maximum-Likelihood Trees for Large Alignments. PLoS ONE, 5(3):e9490. doi:10.1371/journal.pone.0009490.

Data Availability

We gratefully acknowledge all data contributors, i.e. the Authors and their Originating laboratories responsible for obtaining the specimens, and their Submitting laboratories for generating the genetic sequence and metadata and sharing via the GISAID Initiative, on which this research is based. Results for the *in silico* assessment of antigenic lineage alterations that were used to create the heatmaps and global antigenic maps for the months of November 2021 through January 2022 and January 2023 to March 2023 can be found on Zenodo (https://zenodo.org/records/10171227).

Acknowledgements

A.C.M. gratefully acknowledges funding from the German Centre for Infection Research project TI 12.002.

Author Information

Authors and Affiliations

Computational Biology of Infection Research, Helmholtz Centre for Infection Research, Braunschweig, Germany

K. Norwood, Z.-L. Deng, S. Reimering, G. Robertson, M.-H. Foroughmand-Araabi & S. Goliaei

Braunschweig Integrated Centre of Systems Biology (BRICS), Technische Universität Braunschweig, Braunschweig, Germany

K. Norwood, M.-H. Foroughmand-Araabi, S. Goliaei & A. C. McHardy

Genome Competence Center (MF1), Robert Koch Institute, Berlin, Germany

M. Hölzer

Deutsches Zentrum für Infektionsforschung (DZIF), Braunschweig, Germany

A. C. McHardy

Contributions

K.N, Z-L.D, G.R, S.G., M.H. and S.R developed the methods and wrote the code. K.N, S.R. and A.C.M analysed and interpreted the data, K.N, S.R. and A.C.M. wrote the paper and A.C.M. designed the research study. M.H. suggested data visualizations and commented on the manuscript, K.N, Z-L.D, G.R. and M.H.F.A and S.G maintain the web service. All authors read and approved the manuscript.

Corresponding Authors

Correspondence to A. C. McHardy (amc14@helmholtz-hzi.de)

Ethics Declaration

Competing interests

The authors declare no competing interests.