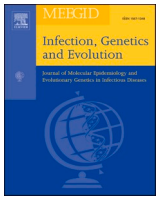
Infection, Genetics and Evolution 119 (2024) 105577

Contents lists available at ScienceDirect

Infection, Genetics and Evolution

journal homepage: www.elsevier.com/locate/meegid



Evolution of SARS-CoV-2 in the Rhine-Neckar/Heidelberg Region 01/2021 – 07/2023

Christian Bundschuh [a,*], Niklas Weidner [a,b], Julian Klein [a], Tobias Rausch [c], Nayara Azevedo [c], Anja Telzerow [a,c], Jan-Philipp Mallm [d], Heeyoung Kim [e], Simon Steiger [f], Isabelle Seufert [f], Kathleen Börner [a], Katharina Bauer [d], Daniel Hübschmann [g], Katharina Laurence Jost [a], Sylvia Parthé [a], Paul Schnitzler [a], Michael Boutros [h], Karsten Rippe [f], Barbara Müller [a], Ralf Bartenschlager [d,i,j], Hans-Georg Kräusslich [a,j], Vladimir Benes [c,*]

[a] Heidelberg University, Medical Faculty Heidelberg, Department of Infectious Diseases, Virology, Heidelberg, Germany
[b] Heidelberg University, Medical Faculty Heidelberg, Department of Infectious Diseases, Microbiology and Hygiene, Heidelberg, Germany
[c] Genomics Core Facility, European Molecular Biology Laboratory (EMBL), Heidelberg, Germany
[d] Single Cell Open Lab, German Cancer Research Center (DKFZ) and Bioquant, Heidelberg, Germany
[e] Department of Infectious Diseases, Molecular Virology, Heidelberg University, Heidelberg, Germany
[f] Division Chromatin Networks, German Cancer Research Center (DKFZ) and Bioquant, Heidelberg, Germany
[g] Division Stem Cells and Cancer, German Cancer Research Center (DKFZ), Heidelberg, Germany
[h] Division Signaling and Functional Genomics, German Cancer Research Center (DKFZ), Department for Cell and Molecular Biology, Medical Faculty Mannheim, Heidelberg University, Heidelberg, Germany
[i] Division Virus-Associated Carcinogenesis, German Cancer Research Center, Heidelberg, Germany
[j] Deutsches Zentrum für Infektionsforschung, partner site Heidelberg, Germany

ARTICLE INFO

Keywords:
SARS-CoV-2
Epidemiology
Whole genome sequencing
Heidelberg/Rhine Neckar region
ARTIC
Bioinformatics

ABSTRACT

In January 2021, the monitoring of circulating variants of SARS-CoV-2 was initiated in Germany under the Corona Surveillance Act, which was discontinued after July 2023. This initiative aimed to enhance pandemic containment, as specific amino acid changes, particularly in the spike protein, were associated with increased transmission and reduced vaccine efficacy.

Our group conducted whole genome sequencing using the ARTIC protocol (currently V4) on Illumina's NextSeq 500 platform (and, starting in May 2023, on the MiSeq DX platform) for SARS-CoV-2 positive specimen from patients at Heidelberg University Hospital, associated hospitals, and the public health office in the Rhine-Neckar/Heidelberg region. In total, we sequenced 26,795 SARS-CoV-2-positive samples between January 2021 and July 2023. Valid sequences, meeting the requirements for upload to the German electronic sequencing data hub (DESH) operated by the Robert Koch Institute (RKI), were determined for 24,852 samples, and the lineage/clade could be identified for 25,912 samples.

The year 2021 witnessed significant dynamics in the circulating variants in the Rhine-Neckar/Heidelberg region, including A.27.RN, followed by the emergence of B.1.1.7 (Alpha), subsequently displaced by B.1.617.2 (Delta), and the initial occurrences of B.1.1.529 (Omicron). By January 2022, B.1.1.529 had superseded B.1.617.2, dominating with over 90%. The years 2022 and 2023 were then characterized by the dominance of B.1.1.529 and its sublineages, particularly BA.5 and BA.2, and more recently, the emergence of recombinant variants like XBB.1.5.

Since the global dominance of B.1.617.2, the identified variant distribution in our local study, apart from a time delay in the spread of new variants, can be considered largely representative of the global distribution.

om a time delay in the spread of new variants, can be considered largely representative of the global distribution.

* Corresponding authors.
E-mail addresses: christian.bundschuh@med.uni-heidelberg.de (C. Bundschuh), benes@embl.de (V. Benes).

https://doi.org/10.1016/j.meegid.2024.105577
Received 12 October 2023; Received in revised form 31 January 2024; Accepted 21 February 2024
Available online 23 February 2024
1567-1348/© 2024 Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

1. Introduction

Towards the end of 2019, a novel coronavirus (2019-nCoV), later termed severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) by the World Health Organization (WHO), emerged as the causative agent for a surge in pneumonia cases identified as coronavirus disease (COVID-19) in Wuhan, China (World Health Organization (WHO), 2020a). The swift propagation of the virus resulted in the transformation of the epidemic in China into a global pandemic, with over 676 million confirmed SARS-CoV-2 cases by March 2023 (data discontinued thereafter) (Center for Systems Science and Engineering (Johns Hopkins University), 2023, World Health Organization (WHO), 2023).

However, given that only a fraction of acute SARS-CoV-2 infections were officially diagnosed and reported, seroprevalence studies suggest that the actual number of cases could be up to 10 times higher (Clarke et al., 2022; Havers et al., 2020; Stringhini et al., 2020).

The Spike (S) protein (with a size of 180–200 kDa) was identified as the key element in host cell entry. Comprising an extracellular N-terminus, a transmembrane (TM) domain penetrating the viral membrane, and a concise intracellular C-terminus (Bosch et al., 2003), the S protein consists of 1273 amino acids, functionally divisible by their respective domains (Xia et al., 2020).

Significant mutations in the S protein have been linked to increased viral transmissibility, severity of the disease, reinfection despite natural immunity, and vaccine efficacy. Monitoring circulating variants has thus become a pivotal epidemiological strategy for pandemic containment (Abdool Karim and de Oliveira, 2021).

Particularly noteworthy are the variants of concern (VOC), which exhibited epidemiological advantages, necessitating the tracking of circulating SARS-CoV-2 variants. Consequently, viral genome sequencing has become indispensable for monitoring the emergence of mutations defining existing and new VOC.

This report details the outcomes of our group's sequencing endeavors in the Rhine-Neckar/Heidelberg region spanning from January 2021 to the close of July 2023. We present findings on the local emergence of a variant labeled as A.27.RN, characterized by the simultaneous presence of the N501Y and L452R mutations in the S gene during the initial quarter of 2021. This variant was subsequently displaced by B.1.1.7, followed by B.1.617.2 and B.1.1.529 (alongside its numerous subvariants, notably the recombinant variants and BA.2.75).

2. Materials and methods

2.1. RNA purification

For SARS-CoV-2 PCR and sequencing analysis, RNA was isolated from upper respiratory tract specimens such as nasopharyngeal and oropharyngeal swabs as well as pharyngeal washes. The isolation was performed by the utilization of automated magnetic bead-based nucleic acid extraction protocols.

Our group utilized either the QIASymphony, DSP Virus/Pathogen mini-Kit (Qiagen) or the Chemagic Viral DNA/RNA 300 Kit H96 (PerkinElmer) for magnetic bead RNA extraction according to the manufacturers' protocols.

2.2. RT-qPCR

Samples were screened for SARS-CoV-2 RNA by commercially available dual-target RT-PCR assays using automated analysis for E & orf1a/b employing Cobas 6800 (Roche Diagnostics) or analysis for E & N Gene (TIB Molbiol) or E & S gene (Altona Diagnostics) by Roche LightCycler and LightCycler II. Positive samples with a Ct value ≤ 35 (reduced to Ct ≤ 32 in week 10 because of a high rate of inconclusive results in samples with Ct > 32) were subsequently characterized by whole genome sequencing. All commercial assays and devices were used according to the manufacturer's instructions.

2.3. SARS-CoV-2 genome sequencing

Following an initial trial period concluding in March 2021, wherein we compared the Nextera library generation approach (Cov-seq) with the ARTIC protocol (Tyson et al., 2020), initially devised by New England Biolabs, our subsequent whole genome sequencing endeavors exclusively employed the ARTIC protocol (currently V4).

The ARTIC protocol implemented in this study closely adheres to the recommended procedures, with three specific optimizations incorporated during the sample upscaling. These enhancements were introduced through the utilization of the NEBNext ARTIC SARS-CoV-2 FS Library Prep Kit for Illumina:

- Omission of the Magnetic Beads Clean-up step after PCR amplification of cDNA.
- The clean-up of adapter-ligated fragments was performed using an automated liquid-handling system.
- After the final PCR, the libraries were pooled and cleaned with SPRIBeads at a 0.9× ratio (bead). All generated libraries were sequenced in pools of up to 384 samples using the Illumina NextSeq 500/MiSeq DX sequencer in paired-end mode with a read length of 75 bases.

Due to its length, only the important key points will be reported in this chapter - the complete, modified ARTIC protocol is shown in the appendix section.

The cDNA synthesis was performed by mixing the RNA sample with the LunaScript RT SuperMix and subsequently running a thermocycler program with primer annealing, synthesis, and heat inactivation steps.

The generated cDNA is amplified in two individual pool set assays (each with their respective ARTIC-SARS-CoV-2 primer mixes) in thermocyclers and then merged together.

Fragmentation (30 min) and end-preparation are performed in the same thermocycler run using the NEBNext Ultra II FS Enzyme Mix, resulting in fragments of approximately 100 base pairs in length.

The adapter ligation is carried out using the NEBNext Adaptor for Illumina and the NEBNext Ultra II Ligation Master Mix. After an incubation phase at 20 °C, (red) USER® enzyme is added, and the mixture is subsequently incubated at 37 °C.

The PCR enrichment of adapter-ligated DNA is performed using the NEBNext Library PCR Master Mix and Index/Universal primers, followed by amplification in a thermocycler run.

Pooling of the libraries can be done without prior quality control, and purification is carried out using NEBNext Sample Purification Beads. The first purification step involves a 0.7-fold ratio of purification beads (based on the pool volume), followed by several incubation, ethanol wash, and drying steps. The final bead extraction is performed using TE buffer and a magnet. The second cleanup step is done with a 0.9-fold ratio of purification beads (based on the pool volume) and otherwise follows the same procedure as the first cleanup.

The last step of the library preparation is the quality control of the final pool using a Qubit DNA BR Kit and Agilent Bioanalyzer High Sensitivity DNA Reagents for sample dilution, the expected peak size is 200–250 base pairs with a 30 min fragmentation time.

2.4. Data analysis

2.4.1. Nextera

Sequencing data from Cov-seq were processed using the nf-core viralrecon pipeline (Ewels et al., 2020, Patel et al., 2020) with the following settings:

- NC_045512.2 genome
- Metagenomic protocol without primer sequence removal
- No duplicate filtering
- Minimal coverage of 20 for variant calling
- Maximum allele frequency of 0.9 for filtering variant calls.

Primer sequences were not removed due to the protocol's specific amplicon design. Variant calls and reconstructed consensus sequences from iVar were used (Grubaugh et al., 2019).

2.4.2. ARTIC

For the data analysis of the ARTIC sequencing, the samples were always sequenced in paired-end mode with a read-length of 75 bp using the NEBNext ARTIC library preparation kit, the detailed protocol is shown in Appendix B. For the data analysis, we applied a multi-stage data processing pipeline that included various tools for sequencing adapter trimming, alignment, quality control, mutation calling, viral genome assembly and lineage classification, which are listed in the following paragraphs.

- Sequencing adapter trimming

Short input fragments may cause sequencing adapters to be present in the final sequencing read. We applied trim galore with default parameters to trim off these adapters (https://zenodo.org/badge/latestdoi/62039322).

- Host-read contamination removal

Because of unspecific amplicon primer binding, the sequencing data may contain contaminating human derived reads. These host reads were removed using a k-mer based method, called kraken2 (Wood et al., 2019), and using alignments against the host reference (build GRCh38). We also applied kraken2 to screen for other potential lab contaminants (Wood et al., 2019).

- Alignment to SARS-CoV-2 reference genome

All remaining sequencing reads were aligned to the SARS-CoV-2 reference genome (NC_045512.2) using the Burrows-Wheeler Alignment tool (Li and Durbin, 2009). Alignments were sorted and indexed using SAMtools (Li et al., 2009) and quality-controlled using Alfred (Rausch et al., 2019). We masked priming regions with iVar (Grubaugh et al., 2019) to avoid a variant calling bias in such loci. Due to the overlapping amplicon design, this strategy does not cause further dropout (unobserved) regions.

- Variant calling

Variant calling employed FreeBayes (Garrison and Marth, 2012) and bcftools (Li, 2011) for quality-filtering and normalization of variants. All variants were annotated for their functional consequence using the Ensembl Variant Effect Predictor (McLaren et al., 2016).

- Consensus computation

SAMtools, bcftools, iVar and Alfred were also employed for viral consensus sequence generation and its quality control. The iVar consensus parameters and quality control stringency were based on the Robert Koch Institute (RKI) SARS-CoV-2 sequencing submission criteria, namely a 90% informative sequence, $<$ 5% Ns (unobserved bases), a minimum of 20$\times$ coverage and at least 90% read support for informative positions (Robert Koch Institute (RKI), 2021b).

- Lineage classification

Lineage and clade classification was carried out using Pangolin (Rambaut et al., 2020) and Nextclade (Hadfield et al., 2018), respectively, with default parameters. At the beginning of the project, where multiple wild-type lineages were present, we also employed a simplified typing system that only highlighted the variants of concern (at the beginning of our project: B.1.1.7, B.1.351 and P.1).

- Summary report generation

We applied a custom python script available in the GitHub source code repository (https://github.com/tobiasrausch/covid19) to aggregate all QC metrics, lineage labels and variants leading to amino acid changes. This report generation also entailed the computation of metadata tracking sheets and gzipped FASTA files for all viral assemblies to facilitate an immediate upload to the German electronic sequencing data hub (DESH) operated by the RKI.

In rare cases, we validated mutations by PCR and Sanger sequencing to confirm interesting sequencing results with the help of primer design and Sanger sequencing chromatogram analysis methods (Rausch et al., 2020) available at GEAR genomics.

2.5. Quality criteria for sequencing data

Due to governmental regulations outlined in the Coronavirus-Surveillanceverordnung (CorSurV) by Germany and Baden-Württemberg, SARS-CoV-2 viral genome sequences were submitted to the DESH database of the Robert Koch Institute (RKI) (Robert Koch Institute (RKI), 2021a, Robert Koch Institute (RKI), 2021b). The sequencing quality of genomes analyzed using both the Nextera and ARTIC protocols was contingent upon meeting the quality criteria established by the RKI. Table 1 (Robert Koch Institute (RKI), 2021b) delineates the key benchmarks constituting the RKI criteria for Illumina-based SARS-CoV-2 genome sequencing.

Further information regarding the respective quality criteria can be found on the RKI webpage (Robert Koch Institute (RKI), 2021b).

2.6. Data availability

SARS-CoV-2 viral genome sequences have been uploaded to the DESH database (https://www.rki.de/DE/Content/InfAZ/N/Neuartiges_Coronavirus/DESH/DESH.html) of the Robert Koch Institute and will be made available via the GISAID database (Shu and McCauley, 2017).

2.7. Ethical statement

This study was approved by the ethics committee of the Medical Faculty at the University of Heidelberg for the analysis of proband samples by whole genome sequencing of the viral RNA (S-316/2021) as well as the virus propagation and neutralization assays and use of sera from vaccinated donors (S-203/2021). Furthermore, the sequencing of SARS-CoV-2 positive samples adhered to the governmental regulations outlined in the Coronavirus-Surveillanceverordnung (CorSurV) by Germany and Baden-Württemberg.

3. Results

3.1. Method comparison Nextera – Artic

Initially, our research team conducted a comparative analysis of two distinct Next-Generation Sequencing (NGS) protocols for SARS-CoV-2

Table 1
RKI quality criteria for DESH upload.

Quality criterion	Threshold
Length of acceptable reads	$\geq$ 30 bp
Average quality of acceptable reads (PHRED)	$\geq$ 20
Sequence identity with NC_045512.2	$\geq$ 90%
Non-aligned bases in reconstructed genome	$\leq$ 5%
Minimal local read depth without PCR duplicate filtering	20
Minimal local read depth with PCR duplicate filtering	10
Informative allele frequency	$\geq$ 90%
Frameshift mutations	Special coverage

sequencing, namely the Nextera and ARTIC protocols. The evaluation was carried out in accordance with the quality criteria specified by the Robert Koch Institute (RKI), and detailed information regarding these protocols is expounded upon in the Materials and Methods section.

These comparative assessments involved 1,488 RNA specimens extracted from the Department of Virology at the Centre of Infectious Diseases, Heidelberg, and the Institute for Medical Microbiology and Hygiene at Mannheim University Hospital. The samples were procured during the months of January and February 2021.

Initial analysis was conducted on a total of 1055 specimens with a positive SARS-CoV-2 PCR result without a Ct cutoff value. As previously mentioned, the benchmark for the required quality in viral genome sequencing was established in accordance with the criteria outlined by the Robert Koch Institute (RKI). The specific outcomes of these analyses are outlined in Table 2.

The ARTIC protocol exhibited markedly superior sequencing results, as per the RKI criteria, compared to the Nextera protocol before the introduction of a Ct cutoff value (*P*-value <0.0001). Subsequently, samples with Ct values > 32 were excluded from further analysis due to elevated failure rates. The introduction of a Ct cutoff value was then validated through testing with 433 specimens, the results are presented in Table 3.

Following the implementation of the Ct cutoff, notable improvements were observed in the performance of both protocols. For the Nextera protocol, the pass rate surged from 3.7 to 45.5% (*P*-value <0.0001), signifying a substantial enhancement. Similarly, the ARTIC protocol demonstrated a pass rate of 91.5% compared to the earlier 68.1% (P-value <0.0001), indicating a significant improvement in sequencing success.

It's noteworthy that incorrect detection of viral genomes was only identified in isolated cases for both protocols, underscoring the overall reliability of the sequencing outcomes.

- Nextera
  - 2 mutated variants identified as wild-type
  - 1 mutated variant identified as a different variant (B.1.1.7 ➔ A.27.RN)
- ARTIC
  - 1 mutated variant identified as wild-type

To cross-verify NGS results, we correlated them with VirSNiP melting curve assays targeting key genetic markers, including N501Y, DelHV69/70, and K417N (TIBMolbiol, Berlin, Germany).

Given the sustained and significant superiority of the ARTIC protocol over the Nextera protocol, both before and after the introduction of a Ct threshold (P-value <0.0001), all subsequent sequencing runs were exclusively conducted utilizing the ARTIC protocol. This protocol was implemented on a NextSeq 500/MiSeq DX platform by Illumina.

3.2. SARS-CoV-2 evolution in 2021

Between January and May 2021, a SARS-CoV-2 variant featuring the L452R and N501Y mutations was identified in the Rhine-Neckar region, detected in 176 out of 3130 specimens. The peak incidence occurred from January to March 2021, with 166 out of 1543 samples acquired during this period. This variant, assigned to the phylogenetic lineage A.27.RN, first surfaced in the Rhine-Neckar/Heidelberg region in a SARS-CoV-2 positive specimen collected during calendar week (CW) 01/2021.

Table 2
Results of the NGS runs before Ct cutoff.

NGS protocol	Pass	Pass [percent]	Fail	Fail [percent]
Nextera	39	3.7	1016	96.3
ARTIC	718	68.1	337	31.9

Table 3
Results of the NGS runs after Ct cutoff.

NGS protocol	Pass	Pass [percent]	Fail	Fail [percent]
Nextera	197	45.5	236	54.5
ARTIC	396	91.5	37	8.5

The defining S gene mutations characterizing the A.27.RN variant were:

- L18F
- L452R
- N501Y
- A653V
- H655Y
- D796Y and
- G1219V

in a D614 wild-type background. The detection patterns of both B.1.1.7 and A.27.RN showed a similar upward trend until CW 7, after which they diverged. Given the prior characterization of A.27.RN in our group's earlier publication (Mallm et al., 2021), it won't be discussed any further at this point.

From CW 5 onward, our results indicated a notable increase of B.1.1.7 until Q2/2021. The prevalence rates of B.1.351, as well as P.3, remained consistently below the 5% threshold in Q1/2021 (CW 1 and 13 exhibited B.1.351 incidence rates slightly below 10%).

In Q2/2021, the rates of B.1.351 stayed below the 5% threshold until week 19, with only sporadic occurrences afterward and no further instances beyond the end of Q2/2021.

Although initially perceived as the most concerning among the early 2021 variants, P.1 demonstrated low incidence numbers in the Rhine-Neckar/Heidelberg region and never reached epidemiological relevance. P.1 was infrequently identified in specimens for a span of 6 weeks before its disappearance by the end of Q2/2021, mirroring the occurrences of B.1.351.

The presence of B.1.617.1 (Kappa) was limited to a solitary specimen in week 14 from the Rhine-Neckar region, with no occurrences in specimens from the Mannheim region.

In Q2/2021, the supremacy of B.1.1.7 was rapidly supplanted by the prevalence of B.1.617.2, with its initial detection in CW 19 and reaching a 50% share by CW 26. By the onset of Q3/2021, all other variants were practically "extinct," emerging only sporadically in resequencing cases of long-term SARS-CoV-2-PCR-positive patients. The dominance of B.1.617.2 persisted until the first week of December.

At that time, the first B.1.1.529 cases were detected in specimens from returning travelers, marking a shift in the prevailing variants. After the first two incidences, within a short timeframe (similar to the dynamics observed for B.1.1.7 and subsequently B.1.617.2), the incidence rate of B.1.1.529 rose to over 40% by the end of 2021.

The initial B.1.1.529 lineage was subsequently reclassified as BA.1.

3.3. Overview of variant distribution in 2021

In 2021, our research group conducted whole genome sequencing on 7995 SARS-CoV-2 positive specimens from the Rhine-Neckar/Heidelberg region. Lineage and clade information could be determined for 7456 samples (93.3%), with only 539 specimens remaining unidentified (6.7%). Among these, 7137 samples met the minimal sequencing criteria set by the RKI DESH (89.3%), while 858 specimens did not meet these criteria (10.7%).

The most significant gap in specimens, for which a sequence could be determined but did not meet the RKI DESH criteria, was observed in the wild-type group. This discrepancy can be attributed to the fact that the highest number of wild-type samples was sequenced before the implementation of the Ct value-dependent sample inclusion criterion (Ct ≤

32). Consequently, more degraded RNA samples were sequenced during this period, leading to an overall higher failure rate in meeting the RKI DESH criteria.

Fig. 1 provides an overview of the number of (successfully) sequenced samples per CW in 2021.

Fig. 2 offers a comprehensive overview of the variant distribution in the Rhein-Neckar/Heidelberg region in 2021.

Detailed tabular overviews of the sequencing results for 2021, 2022, and 2023, as well as the variant classifications for wild type and recombinant variants, can be found in Appendix A.

3.4. SARS-CoV-2 evolution in 2022

In 2022, B.1.1.529 (BA.1) swiftly became the predominant variant in the Rhein-Neckar/Heidelberg region, almost entirely displacing B.1.617.2 by the end of January. However, the ascendancy of BA.1 was brief. Before a full displacement of B.1.617.2 occurred, BA.1's prevalence waned, and the emergence of BA.2 commenced. Within a span of two months, until the end of March 2022, BA.1's prevalence dipped below the 10% threshold. Nevertheless, BA.2 was almost displaced by subsequent variants until the emergence of recombinant Omicron variants, which currently primarily comprise recombinations among BA.2 sublineages.

Similar to global trends, BA.3 struggled to establish itself in the Rhein-Neckar/Heidelberg region, with only isolated cases reported in CW 13. Although BA.4 and BA.5 were identified nearly simultaneously between CW 15 and 19, BA.4 failed to achieve epidemiologically significant numbers. Despite not being entirely displaced by BA.5, BA.4 never reached a proportion of 10% of total infections per CW. In contrast, once BA.5 surpassed the 10% mark, it took just under four weeks to account for over 80% of total infections. Throughout most of the second half of 2022, the proportion of BA.5 remained above the 90% mark. This trend shifted only from CW 45 onward when there was a resurgence of recombinant Omicron variants, such as XBB (which emerged through recombination of BA.2.10.1 and BA.2.75), which subsequently became even more dominant in 2023.

3.5. Overview of variant distribution in 2022

In 2022, our research group conducted sequencing on 17,455 SARS-CoV-2 positive specimens from the Rhine-Neckar/Heidelberg region. Lineage and clade information was determinable for 17,215 samples (98.6%), with only 240 specimens remaining unidentifiable (1.4%). Among these, 16,512 samples met the minimal sequencing criteria set by the RKI DESH (94.6%), leaving 943 specimens that did not (5.4%).

Notably, in 2022, no significant negative gap was identified in sequenced samples that did not meet the RKI DESH criteria, unlike in the year 2021.

An overview of the number of sequenced samples per CW and the success rate of sequencing is depicted in Fig. 3.

Fig. 4 offers a comprehensive overview of the variant distribution in the Rhein-Neckar/Heidelberg region in 2022.

3.6. SARS-CoV-2 evolution in 2023 (until July)

The year 2023 commenced much like the conclusion of 2022, witnessing a rise in the number of recombinant variants, while BA.2(.75) maintained a consistent presence, and cases of BA.5 continued to decline. These patterns persisted until CW 18, marking the final identification of a non-recombinant BA.5. From CW 20 onwards, only recombinant versions were observed, with an exception in CW 26, where two specimens with BA.2.75 were noted.

The dwindling numbers of SARS-CoV-2 positive PCR specimens since the beginning of 2023 (retrospectively evident from CW 42/2022) reached their lowest point from CW 17 onwards, with < 20 specimens per week until the conclusion of the surveillance period at the end of July 2023. Responding to this trend, our group initiated the transfer of the ARTIC protocol from the NextSeq 500 to the MiSeqDX platform at the end of March 2023.

On May 5, 2023, the WHO declared the end of the SARS-CoV-2 pandemic after > 3 years. Subsequently, by the end of May 2023, the DESH platform (Robert Koch Institute (RKI), 2021a) was officially discontinued.

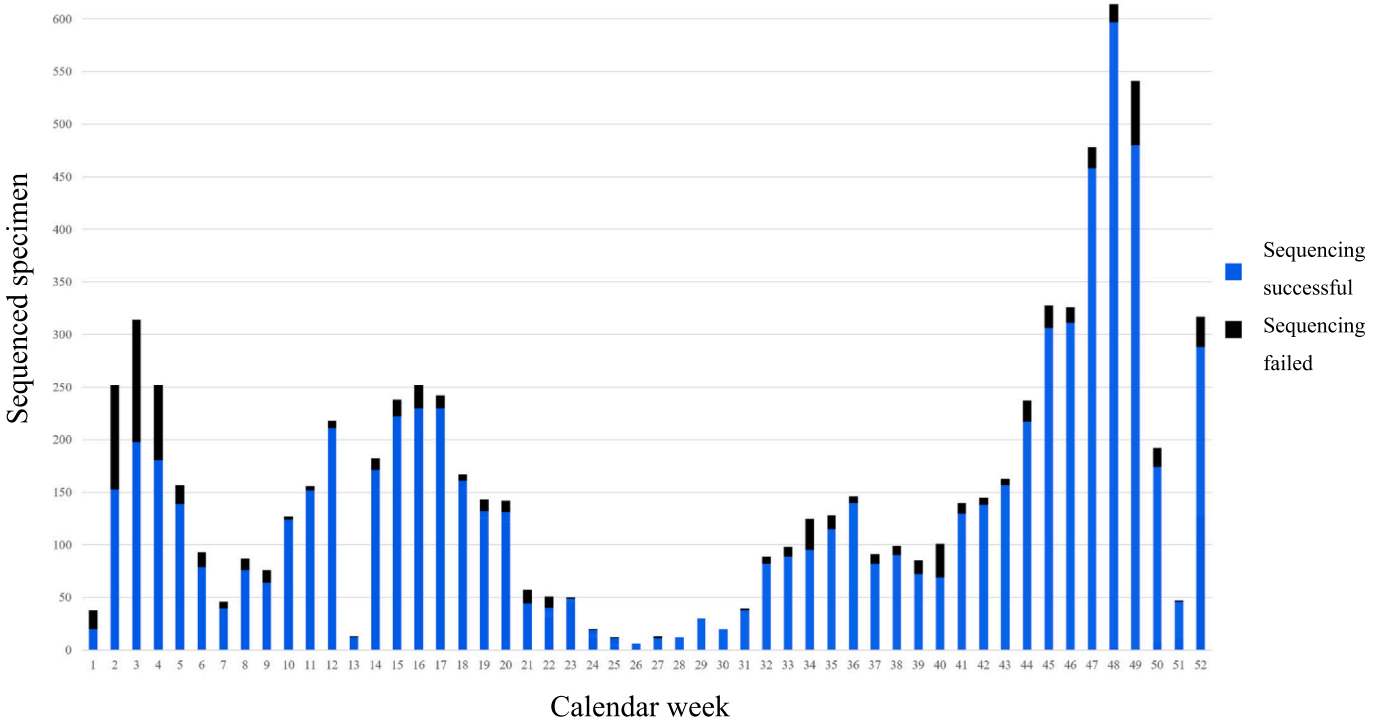


Fig. 1. Overview of successful and failed sequencings for Rhine-Neckar region per calendar week in 2021.

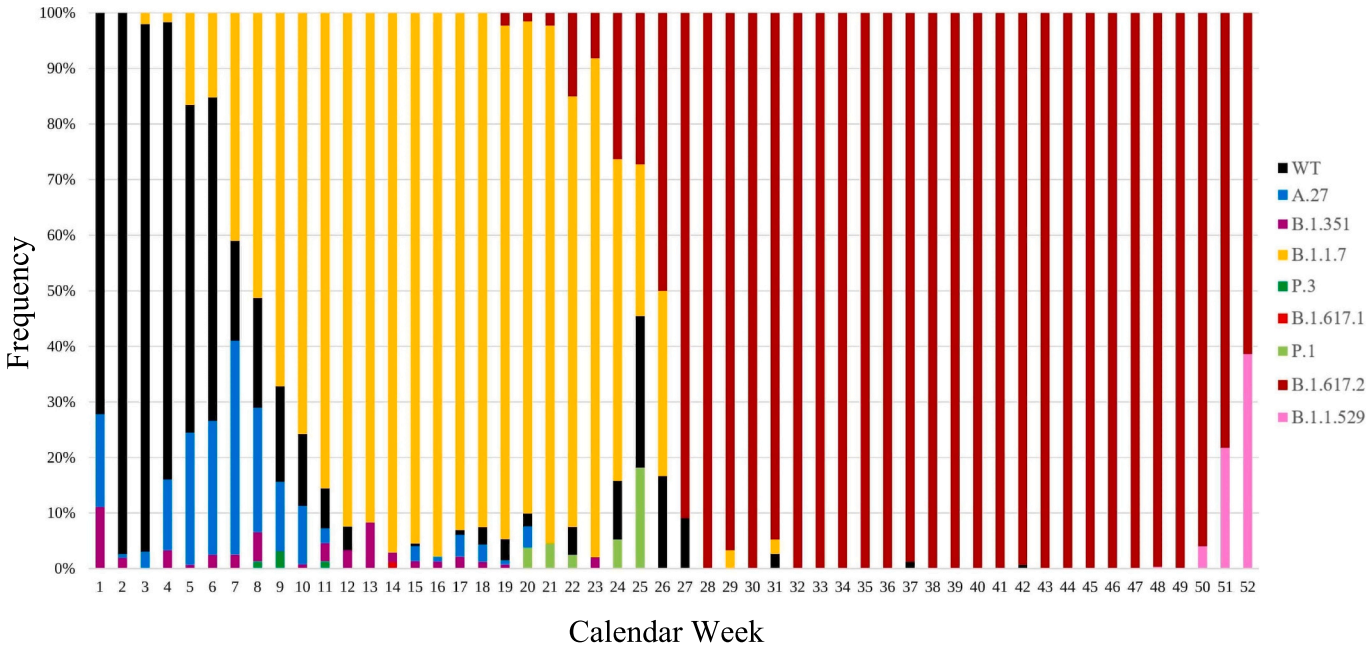


Fig. 2. SARS-CoV-2 variant distribution per calendar week in 2021.

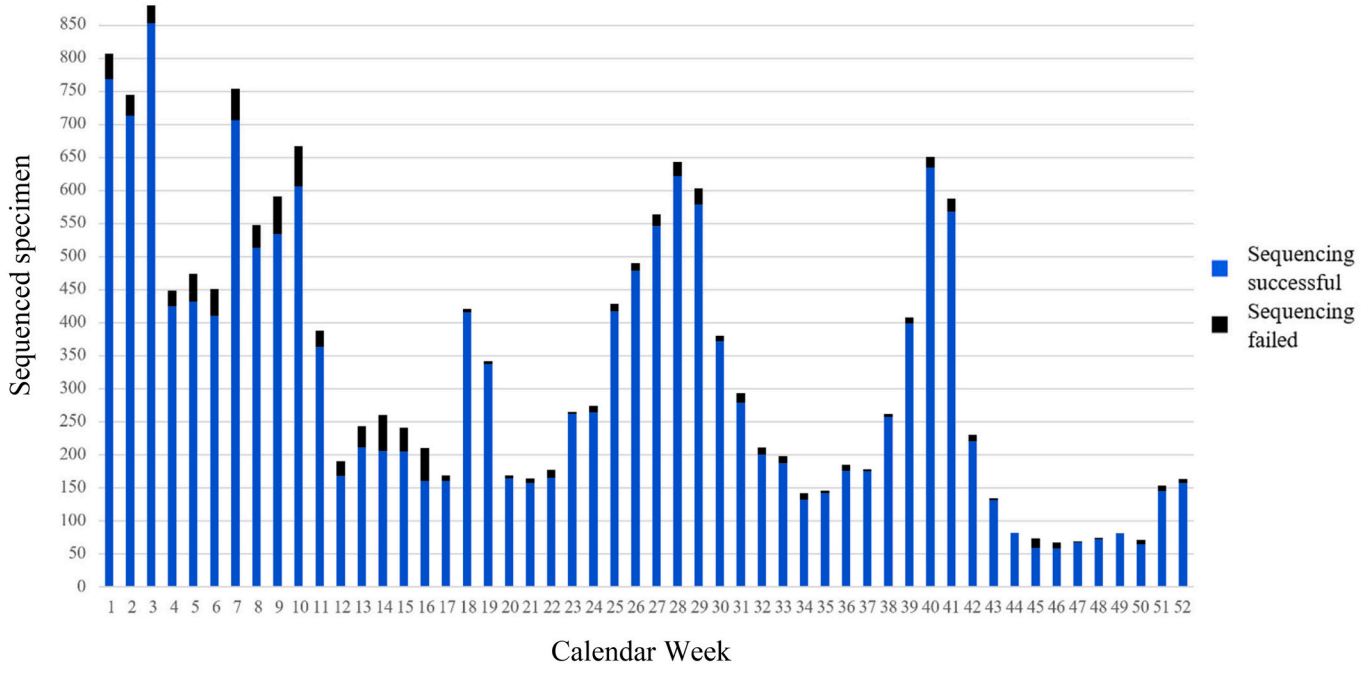


Fig. 3. Overview of successful and failed sequencings for Rhine-Neckar region per calendar week in 2022.

3.7. Overview of variant distribution in 2023 (until July)

In 2023, our research group conducted sequencing on 1345 SARS-CoV-2 positive specimens from the Rhine-Neckar/Heidelberg region. Lineage and clade information was determinable for 1306 samples (97.1%), with only 39 specimens remaining unidentifiable (2.9%). Among these, 1199 samples met the minimal sequencing criteria set by the RKI DESH (89.1%), leaving 146 specimens that did not (10.9%).

An overview of the number of sequenced samples per CW and the success rate of sequencing is depicted in Fig. 5.

Fig. 6 offers a comprehensive overview of the variant distribution in the Rhein-Neckar/Heidelberg region in 2023.

3.8. Phylogenetic relationships

Currently, > 20 main clades related to SARS-CoV-2 are defined by Nextstrain criteria. Notably, 22E, 22F/23A, and the recombinant clades are currently considered epidemiologically relevant due to their even more significant replication advantage compared to previously circulating variants (Aksamentov et al., 2021; Hadfield et al., 2018).

For the reader's convenience (and given its availability under a CC-BY-4.0 license) Fig. 7 provides an overview of the phylogenetic classification of SARS-CoV-2 sorted by clade membership, along with corresponding variants. A continuously updated version of this overview, as well as the toolset used for generating the phylogenetic trees, can be

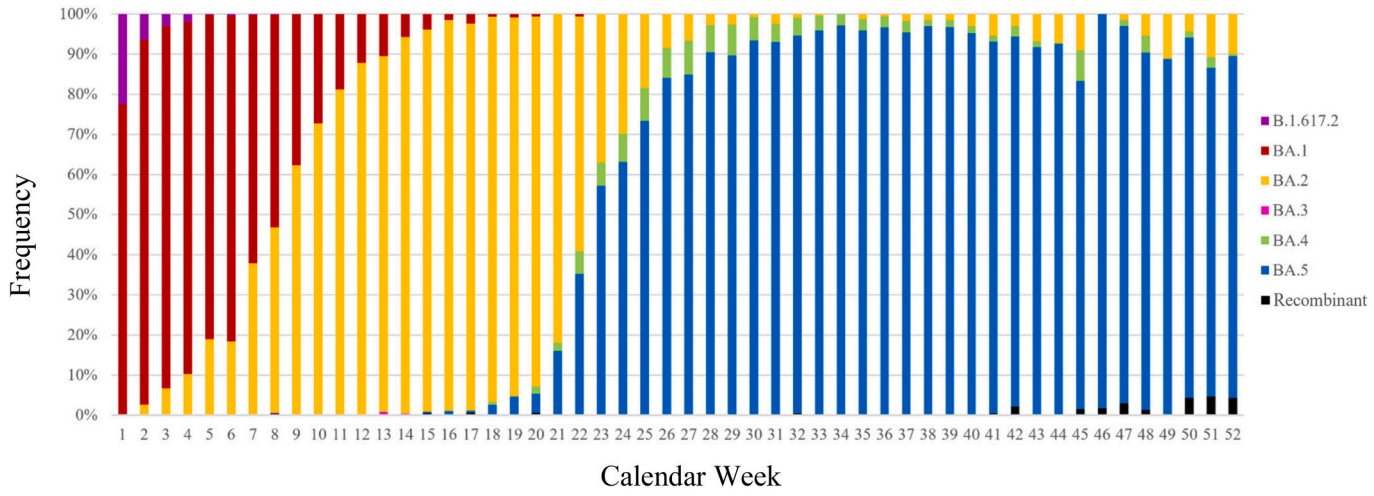


Fig. 4. SARS-CoV-2 variant distribution per calendar week in 2022.

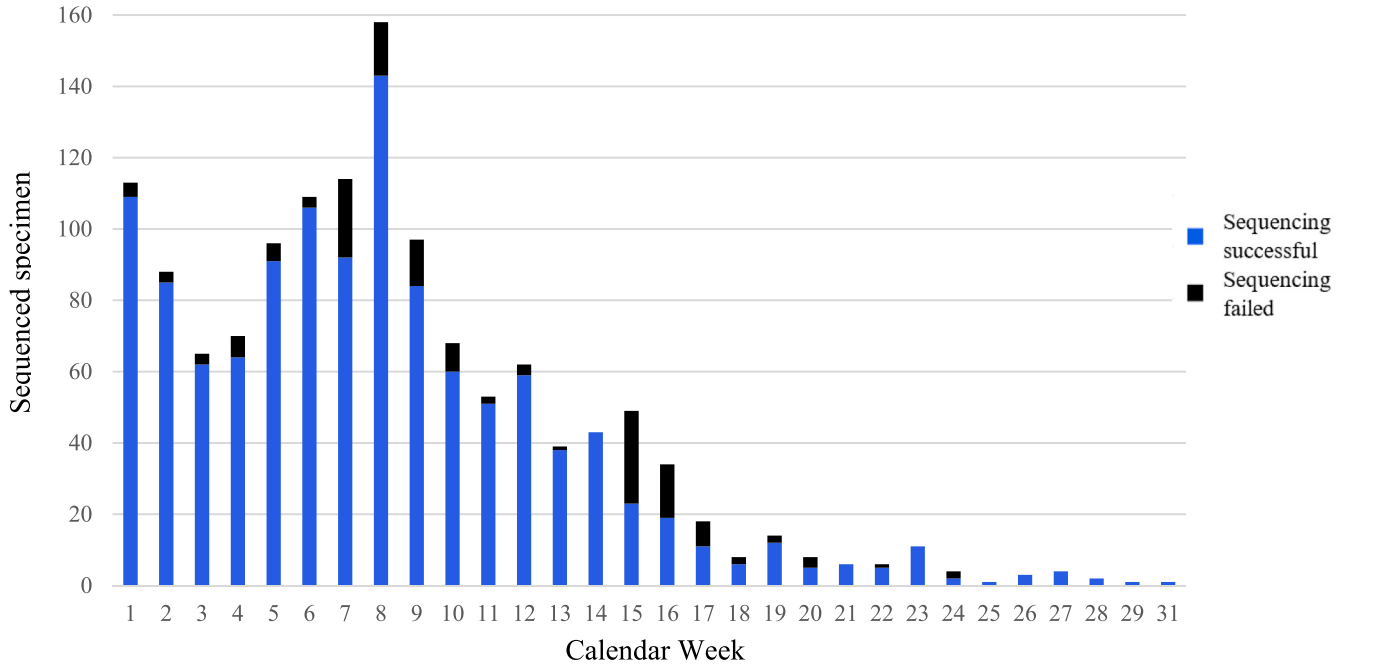


Fig. 5. Overview of successful and failed sequencings for Rhine-Neckar region per calendar week in 2023.

accessed on the Nextclade website and the respective GitHub repository (Aksamentov et al., 2021, Hadfield et al., 2018).

Fig. 8 illustrates the phylogenetic development of SARS-CoV-2 in the Rhine-Neckar region throughout the surveillance project period (01/2021–07/2023). The phylogenetic tree was generated from the collected sequencing data using the Nextclade tool. Due to computational limitations (a maximum of 3,500 sequences in one analysis/generation cycle of the phylogenetic tree), only every eighth sample from the Rhein-Neckar/Heidelberg region, sorted by entry time in the laboratory, could be included in creating the subsequent phylogenetic tree. To prevent possible omissions of low-frequency variants in the phylogenetic overview, care was taken to include every variant detected over the surveillance project's timeframe in the dataset for the Nextclade tool.

4. Discussion

In summary, throughout this project we sequenced a wide variety of SARS-CoV-2 variants with significant differences in their epidemiological relevance. While some variants were restricted to specific geographical regions (Mallm et al., 2021), others, such as B.1.617.2 and B.1.1.529, emerged worldwide.

In the global context, the first half of 2021 (as well as the entirety of 2020) witnessed notable regional disparities in the evolution and dissemination of SARS-CoV-2 variants (Mallm et al., 2021; Urhan and Abeel, 2021). Initial Variants of Concern (VOCs) such as Alpha (B.1.1.7), Beta (B.1.351), and Gamma (P.1) exhibited pronounced regional variations. For instance, Gamma predominated in Brazil, while in our investigations (and in Europe), B.1.1.7 was the prevailing variant. Consequently, during this period, extrapolating results from locally confined studies is reliably possible only within certain limits (extrapolation of our study's results only for Europe).

However, the advent of Delta (B.1.617.2) marked a significant shift due to its global dissemination, allowing for the extrapolation of our locally conducted study's results to larger geographical areas and even globally. Regarding the temporal aspect, there was consistently a delay

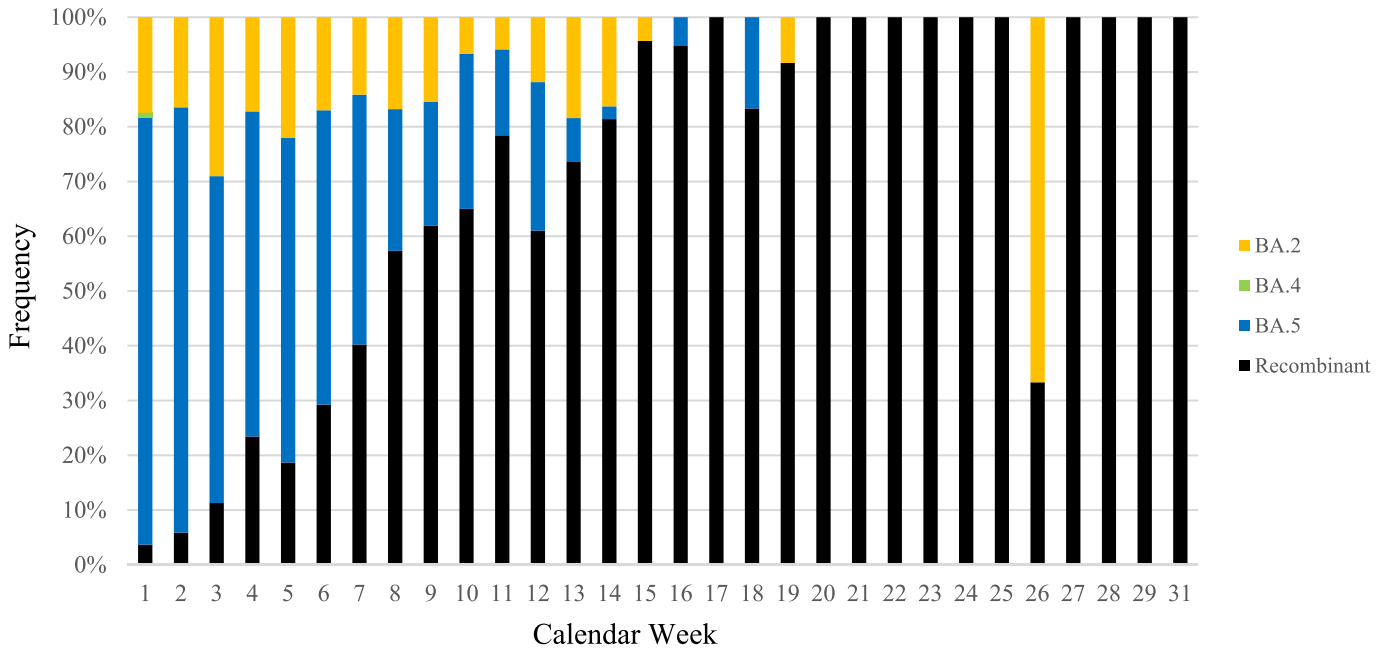


Fig. 6. SARS-CoV-2 variant distribution per calendar week in 2023.

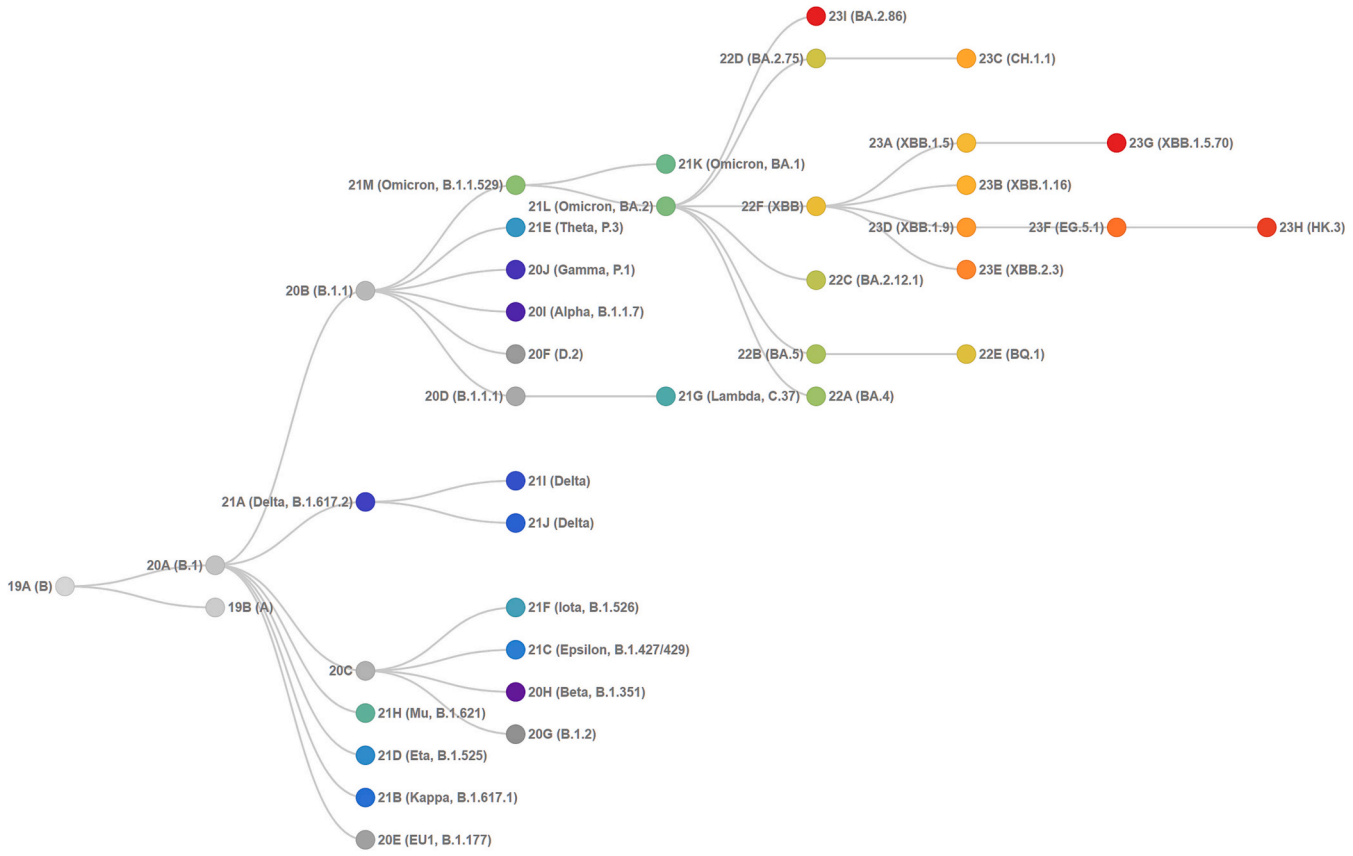


Fig. 7. Phylogenetic classification of SARS-CoV-2 (Aksamentov et al., 2021, Hadfield et al., 2018).

compared to the global initial detection. The time delay for Alpha (B.1.1.7) was at most 3–4 months (September 2020 in the UK, CW 03/2021 in Heidelberg/Rhine Neckar region), for Beta (B.1.351), it was at most 6–7 months (May 2020 in South Africa, CW 01/2021 Heidelberg/Rhine Neckar region), while for Gamma (P.1) (November 2020 in Brazil, CW 20/2021 in Heidelberg/Rhine Neckar region) and Delta (B.1.617.2), it was 5–6 months (October 2020 in India, CW 19/2021 in Heidelberg/Rhine Neckar region) (World Health Organization (WHO), 2020b). Since SARS-CoV-2 positive samples were not sequenced before January 2021 in our study, the possibility of an earlier initial detection cannot be definitively ruled out - especially for Alpha (B.1.1.7) and Beta (B.1.351), due to the nationwide first detection of Beta (B.1.351) in December 2020

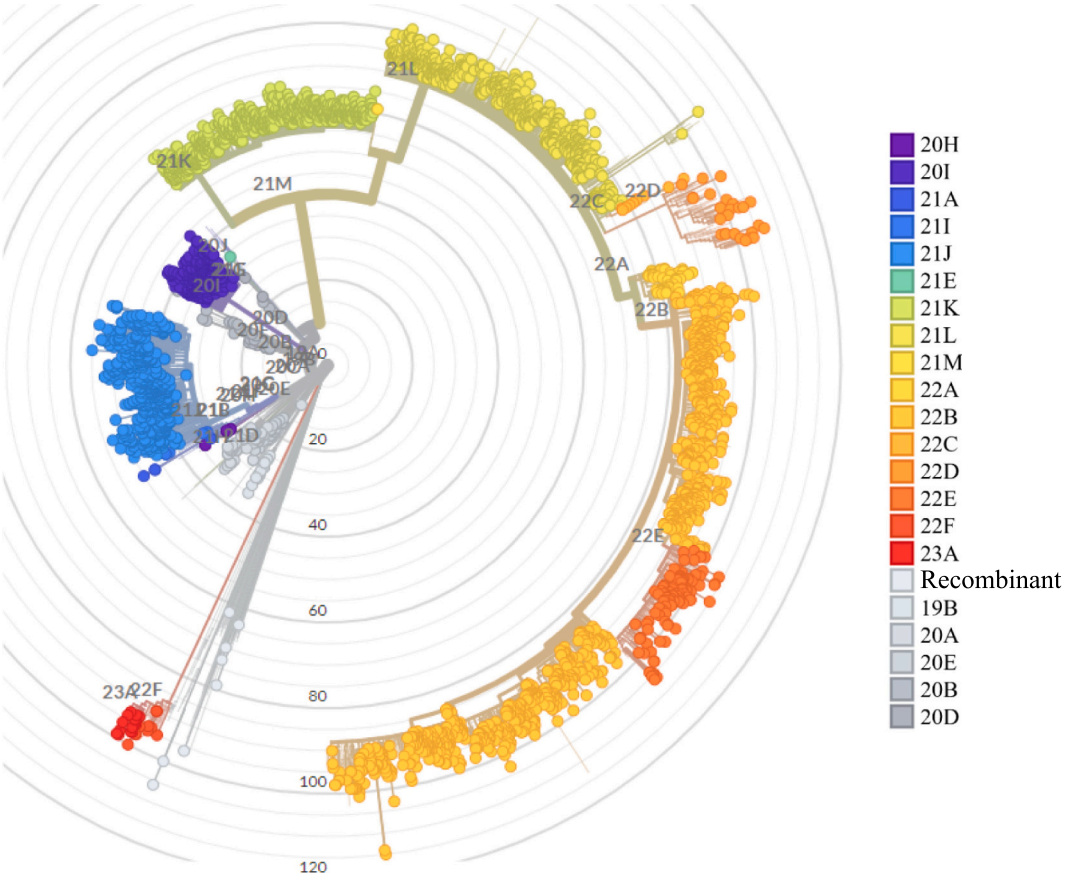


Fig. 8. Phylogenetic tree of SARS-CoV-2 clades in the Rhine-Neckar region 01/2021–07/2023 Generated with the Nextclade Tool.

and of Alpha (B.1.1.7) in the first half of January (Ministerium für Soziales, 2021, Robert Koch Institute (RKI), 2021c).

While these four VOC variants exhibited similar dynamics in their spread/time delay (with the shorter time delay for the UK attributed to the significantly reduced distance compared to the initial detection locations of the other VOCs), Omicron (B.1.1.529) demonstrated a notably shorter time delay of only a few weeks. It was initially detected in two returnees from South Africa in CW 50/2021 in the Heidelberg/Rhine Neckar region (November 2021 in Botswana and South Africa) (World Health Organization (WHO), 2020b). This shorter delay is primarily due to the increased transmission rate and improved immune escape of B.1.1.529 compared to preceding variants (Baker et al., 2022; Jørgensen et al., 2022; Yue et al., 2023).

In summary, it can be inferred that for global pandemic variants (such as Delta (B.1.617.2) and Omicron (B.1.1.529)), extrapolating results from locally confined studies is feasible for assessing the worldwide epidemiological situation. Nonetheless, in the event of a new variant, a time delay should always be considered. Evaluating this time delay also appears to provide an epidemiological understanding of the transmission rate (and immune escape, particularly in the context of reduced vaccine efficacy).

5. Conclusion

It is likely that the current state of B.1.1.529 with its currently circulating recombinant variants is not the final state regarding circulating SARS-CoV-2 variants (or mutations) with significant selective advantages over established variants. Therefore, new variants with even greater epidemiological impact will most likely emerge.

The ongoing evolution and emergence of SARS-CoV-2 variants poses ongoing challenges to public health efforts, requiring continuous monitoring, research, and adaptability in response to changing viral dynamics (although the significant reduction of symptom severity and reduction of detection levels over the last few months, put the monitoring requirement into a new perspective). Yet, vigilance and preparedness are still crucial regarding prospective new SARS-CoV-2 variants with a currently undetermined hazard rate regarding their severity and transmissibility to mitigate the potential impact of new variants on public health.

Therefore, the continued and comprehensive surveillance of emerging SARS-CoV-2 strains and mutations by unbiased whole genome sequencing remains of utmost importance for monitoring the still present evolutionary changes in SARS-CoV-2 variants (even after the end of the pandemic).

6. Limitations

The primary limitation of this study stems from the fact that our laboratory was not the sole sequencing institute in the Rhein-Neckar/Heidelberg region. This aspect was previously acknowledged in 2021 by conducting a comparative analysis of the variant distribution/sequencing results between this study and those of the Limbach Laboratory Heidelberg. The comparison revealed no significant variation in the circulating variants, with differences of $> 10\%$ in individual variant frequencies for at least two consecutive weeks considered as statistically significant.

Founding sources

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

CRediT authorship contribution statement

Christian Bundschuh: Writing – review & editing, Writing – original draft, Investigation, Data curation. **Niklas Weidner:** Writing – review & editing, Investigation, Data curation. **Julian Klein:** Investigation. **Tobias Rausch:** Writing – review & editing, Formal analysis, Data curation. **Nayara Azevedo:** Investigation. **Anja Telzerow:** Investigation. **Jan-Philipp Mallm:** Writing – review & editing, Investigation, Formal analysis. **Heeyoung Kim:** Investigation, Formal analysis. **Simon Steiger:** Formal analysis. **Isabelle Seufert:** Formal analysis. **Kathleen Börner:** Investigation. **Daniel Hübschmann:** Data curation. **Katharina Laurence Jost:** Writing – review & editing, Investigation. **Sylvia Parthé:** Investigation. **Paul Schnitzler:** Writing – review & editing, Supervision. **Michael Boutros:** Writing – review & editing, Supervision. **Karsten Rippe:** Writing – review & editing, Supervision, Conceptualization. **Barbara Müller:** Writing – review & editing, Investigation. **Ralf Bartenschlager:** Writing – review & editing, Supervision, Conceptualization. **Hans-Georg Kräusslich:** Writing – review & editing, Supervision, Conceptualization. **Vladimir Benes:** Writing – review & editing, Investigation, Formal analysis.

Declaration of competing interest

None declared.

Data availability

Data will be made available on request.

Acknowledgments

We thank Christian Drosten and Barbara Mühlemann from the German National Consultant Laboratory for Coronaviruses for discussion on SARS-CoV-2 variant distribution. We gratefully acknowledge the help of Maria Anders-Össwein with serological analyses and of Marie Bartenschlager with the neutralization assays and thank all diagnostics employees and EMBL GeneCore's NGS protocol-automation team for their support. This work was supported by the program for surveillance and control of SARS-CoV-2 mutations of the state of Baden-Württemberg, the German Federal Research Network Applied Surveillance and Testing (B-FAST) within the Network University Medicine and the DKFZ@fightCOVID initiative.

Appendix A

This section provides a detailed/tabular overview of the sequencing results as well as the variants classified as wild type and recombinant variants

The case number reflects the respective variant's sequence frequencies which fulfilled the RKI DESH criteria (Robert Koch Institute (RKI), 2021b) over the respective time frame. The numbers in brackets reflect the total number of specimens assigned to the respective variant following the sequencing process, encompassing samples that both passed and failed the RKI DESH criteria. The first and last discovery dates indicate the initial and final detection dates of the respective variant. Wild type and recombinant variant tables also comprise the respective variant's clade.

Table A.1 presents the sequencing results of 2021.

Table A.1
SARS-CoV-2 variant distribution in the Rhine-Neckar region 2021.

Variant	Cases	First Discovery	Last Discovery
Wild type	722 (794)	02/01/2021	26/06/2021
A.27	170 (176)	14/01/2021	19/05/2021
B.1.1.7	1908 (1961)	20/01/2021	05/08/2021
thereof Q.1	43 (47)	13/02/2021	11/06/2021
thereof Q.4	4 (4)	24/02/2021	24/02/2021
B.1.351	51 (55)	13/01/2021	08/06/2021
P.1	11 (12)	18/05/2021	25/06/2021
P.3	5 (5)	27/02/2021	20/03/2021
B.1.617.1	2 (2)	11/04/2021	11/04/2021
B.1.617.2	4122 (4304)	10/05/2021	31/12/2021
thereof AY.4.2	63 (65)	06/10/2021	18/12/2021
B.1.1.529/BA.1	146 (147)	03/12/2021	31/12/2021
Total	7137 (7456)		

Table A.2 presents the sequencing results of 2022.

Table A.2
SARS-CoV-2 variant distribution in the Rhine-Neckar region 2022.

Variant	Cases	First Discovery	Last Discovery
B.1.617.2	266 (266)	10/05/2021	22/02/2022
BA.1	4279 (4509)	03/12/2021	02/06/2022
BA.2	4489 (4799)	07/01/2022	31/12/2022
BA.3	5 (6)	22/02/2022	06/04/2022
BA.4	374 (382)	02/05/2022	29/12/2022
BA.5	7064 (7216)	14/04/2022	31/12/2022
Recombinant	35 (37)	22/02/2022	30/12/2022
Total	16,512 (17215)		

Table A.3 presents the sequencing results of 2023.

Table A.3
SARS-CoV-2 variant distribution in the Rhine-Neckar region 2023.

Variant	Cases	First Discovery	Last Discovery
BA.2	182 (197)	01/01/2023	30/06/2023
BA.3	0 (2)	18/04/2023	23/04/2023
BA.4	1 (1)	04/01/2023	04/01/2023
BA.5	482 (508)	01/01/2023	04/05/2023
Recombinant	534 (598)	02/01/2023	31/07/2023
Total	1199 (1306)		

Table A.4 presents the evolutionary omicron timeline.

Table A.4
Omicron Distribution in the Rhein-Neckar/Heidelberg Region.

Variant	Cases	First Discovery	Last Discovery
BA.1	4473 (4714)	03/12/2021	02/09/2022
BA.2	4671 (5000)	07/01/2022	30/06/2023
BA.3	5 (8)	22/02/2022	23/04/2023
BA.4	375 (383)	02/05/2022	04/01/2023
BA.5	7546 (7732)	14/04/2022	04/05/2023
Recombinant	569 (636)	22/02/2022	31/07/2023
Total	17,639 (18473)	03/12/2021	31/07/2023

Table A.5 provides a comprehensive overview of the variants classified as wild-type variants, along with their respective incidences, first and last discoveries as well as their clades.

Table A.5
Non-VOC statistics for the Rhine-Neckar region.

Variant	Cases	First Discovery	Last Discovery	Clade
A.5	0 (1)	29/01/2021	29/01/2021	19B
A.29	2 (7)	31/01/2021	24/04/2021	19B
AA.1	2 (3)	14/01/2021	10/03/2021	20E
B.1	9 (20)	17/01/2021	29/05/2021	20A/20B/20E
B.1.1	24 (32)	12/01/2021	02/05/2021	20B
B.1.1.28	2 (2)	25/01/2021	25/01/2021	20B
B.1.1.39	22 (23)	13/01/2021	24/02/2021	20B
B.1.1.70	13 (14)	12/01/2021	14/02/2021	20B
B.1.1.153	3 (4)	19/01/2021	30/01/2021	20B
B.1.1.161	0 (1)	13/01/2021	13/01/2021	20B
B.1.1.170	27 (28)	08/01/2021	20/03/2021	20B
B.1.1.189	6 (6)	05/01/2021	22/01/2021	20B
B.1.1.232	2 (2)	12/01/2021	20/01/2021	20B
B.1.1.236	13 (14)	13/01/2021	27/01/2021	20B
B.1.1.277	1 (1)	15/01/2021	15/01/2021	20B
B.1.1.294	1 (1)	29/01/2021	29/01/2021	20B
B.1.1.297	24 (28)	04/01/2021	01/03/2021	20B
B.1.1.317	7 (7)	13/01/2021	12/03/2021	20B
B.1.1.318	2 (2)	03/05/2021	03/05/2021	20B
B.1.1.347	1 (1)	19/01/2021	19/01/2021	20B
B.1.1.374	2 (2)	04/02/2021	28/02/2021	20B
B.1.1.385	38 (38)	14/01/2021	25/02/2021	20B
B.1.1.519	9 (9)	01/03/2021	28/03/2021	20B
B.1.2	8 (9)	19/01/2021	15/02/2021	20G
B.1.9	12 (14)	03/01/2021	26/03/2021	20A
B.1.36	4 (4)	21/01/2021	15/02/2021	20A
B.1.160	106 (111)	02/01/2021	22/10/2021	20A
B.1.177	143 (154)	06/01/2021	25/03/2021	20E
B.1.221	76 (83)	05/01/2021	13/04/2021	20A
B.1.222	0 (1)	14/01/2021	14/01/2021	20A
B.1.258	129 (135)	04/01/2021	23/03/2021	20A
B.1.406	4 (4)	23/01/2021	05/02/2021	20A
B.1.474	1 (1)	23/01/2021	23/01/2021	20A
B.1.579	0 (1)	20/01/2021	20/01/2021	20A
B.1.619	6 (7)	03/05/2021	25/06/2021	20A
B.1.621	1 (1)	14/06/2021	14/06/2021	21H
B.55	0 (1)	22/01/2021	22/01/2021	19A
C.35	9 (9)	13/01/2021	08/02/2021	20D
C.36	8 (8)	14/ 01/2021	26/06/2021	20D
C.38	1 (1)	16/05/2021	16/05/2021	20D
G.1	1 (1)	21/01/2021	21/01/2021	20A
R.1	3 (3)	26/02/2021	28/02/2021	20B
Total	722 (794)	02/01/2021	26/06/2021	

Table A.6 provides a comprehensive overview of the variants classified as recombinant variants, along with their respective incidences, first and last discoveries as well as their clades.

Table A.6
Recombinant variant(s) statistics for the Rhine-Neckar region.

Variant	Cases	First Discovery	Last Discovery	Clade
EG.1	4 (5)	13/03/2023	28/05/2023	22F
EG.1.2	1 (1)	10/05/2023	10/05/2023	23D
EG.1.4	0 (1)	14/04/2023	14/04/2023	23A
EG.1.6	0 (1)	05/05/2023	05/05/2023	23D
EG.5.1.3	1 (1)	29/06/2023	29/06/2023	23F
EG.5.2.3	1 (1)	21/06/2023	21/06/2023	23D
EG.10	1 (1)	14/05/2023	14/05/2023	23D
EG.10.1	1 (1)	17/07/2023	17/07/2023	23D
EL.1	1 (1)	06/06/2023	06/06/2023	23A
EU.1.1	19 (20)	09/03/2023	06/06/2023	23A
FL.1.2	1 (1)	20/04/2023	20/04/2023	22F
FL.2	2 (2)	01/06/2023	09/06/2023	23D
FL.2.1	4 (4)	14/04/2023	13/05/2023	22F
FL.3.1	1 (1)	30/04/2023	30/04/2023	22F
FL.4	4 (4)	06/06/2023	14/06/2023	23D
FL.5	1 (1)	29/04/2023	29/04/2023	22F
FL.8	1 (1)	13/05/2023	13/05/2023	23D
FL.18	1 (1)	19/05/2023	19/05/2023	23D
FL.19	1 (1)	08/05/2023	08/05/2023	23D
FL.22	1 (1)	10/07/2023	10/07/2023	23D
FU.2	1 (1)	12/04/2023	12/04/2023	23B
GB.1	1 (1)	18/04/2023	18/04/2023	23A
GN.1.1	2 (2)	05/07/2023	14/07/2023	23A
HZ.2	1 (1)	12/06/2023	12/06/2023	23A
XAF	1 (1)	18/05/2022	18/05/2022	Recombinant
XAN	1 (1)	14/08/2022	14/08/2022	Recombinant
XAY.1.1	4 (4)	18/01/2023	22/02/2023	Recombinant
XAY.1.1.1	3 (3)	21/02/2023	22/04/2023	Recombinant
XAZ	1 (1)	12/10/2022	12/10/2022	Recombinant
XBB	1 (2)	05/03/2023	05/03/2023	22F
XBB.1	14 (15)	12/10/2022	21/02/2023	22F
XBB.1.2	4 (4)	14/12/2022	01/02/2023	22F
XBB.1.4	1 (1)	27/12/2022	27/12/2022	22F
XBB.1.5	250 (279)	21/12/2022	07/07/2023	23A
XBB.1.5.1	6 (6)	01/02/2023	10/05/2023	23A
XBB.1.5.3	7 (7)	20/03/2023	11/04/2023	23A
XBB.1.5.7	9 (11)	05/02/2023	28/03/2023	23A
XBB.1.5.12	5 (5)	10/03/2023	28/04/2023	23A
XBB.1.5.13	18 (20)	10/03/2023	06/06/2023	23A
XBB.1.5.14	0 (1)	31/03/2023	31/03/2023	23A
XBB.1.5.15	1 (1)	01/04/2023	01/04/2023	23A
XBB.1.5.20	3 (3)	08/03/2023	16/03/2023	23A
XBB.1.5.21	3 (3)	07/03/2023	16/03/2023	23A
XBB.1.5.24	7 (8)	24/03/2023	30/05/2023	23A
XBB.1.5.33	1 (1)	16/03/2023	16/03/2023	23A
XBB.1.5.37	3 (4)	27/03/2023	20/04/2023	23A
XBB.1.5.38	1 (1)	24/04/2023	24/04/2023	23A
XBB.1.5.43	1 (1)	24/04/2023	24/04/2023	23A
XBB.1.5.49	0 (1)	14/04/2023	14/04/2023	23A
XBB.1.5.59	1 (1)	04/07/2023	04/07/2023	23A
XBB.1.5.62	0 (1)	13/04/2023	13/04/2023	23A
XBB.1.5.63	5 (11)	12/04/2023	02/06/2023	23A
XBB.1.5.86	1 (1)	22/05/2023	22/05/2023	23A
XBB.1.9	0 (1)	14/02/2023	14/02/2023	22F
XBB.1.9.1	80 (92)	23/01/2023	07/06/2023	22F
XBB.1.9.2	20 (20)	23/01/2023	03/04/2023	22F
XBB.1.11.1	1 (1)	25/04/2023	25/04/2023	22F
XBB.1.13	1 (1)	06/04/2023	06/04/2023	22F
XBB.1.16	4 (4)	29/03/2023	31/07/2023	22F
XBB.1.16.1	1 (2))	02/06/2023	02/06/2023	23B
XBB.1.16.21	1 (1)	09/06/2023	09/06/2023	23B
XBB.1.19.1	1 (1)	24/04/2023	24/04/2023	22F
XBB.1.22.2	4 (4)	08/03/2023	05/04/2023	23A
XBB.2	4 (5)	20/10/2022	06/03/2023	22F
XBB.2.3	1 (1)	27/02/2023	27/02/2023	22F
XBB.2.3.4	0 (1)	11/04/2023	11/04/2023	22F
XBB.2.4	2 (2)	03/02/2023	28/02/2023	22F
XBB.3	2 (2)	19/10/2022	09/11/2022	22F
XBD	2 (2)	21/10/2022	25/11/2022	Recombinant
XBF	19 (19)	19/12/2022	27/04/2023	Recombinant

(continued on next page)

Table A.6 (*continued*)

Variant	Cases	First Discovery	Last Discovery	Clade
XBF.3	1 (1)	22/02/2023	22/02/2023	Recombinant
XBJ	0 (1)	15/11/2022	15/11/2022	Recombinant
XBK	14 (14)	21/12/2022	03/05/2023	Recombinant
XBK.1	1 (1)	24/03/2023	24/03/2023	Recombinant
XBL	2 (2)	30/12/2022	22/02/2023	Recombinant
XG	1 (1)	14/04/2022	14/04/2022	Recombinant
XM	4 (4)	22/02/2022	02/05/2022	Recombinant
Total	569 (636)	22/02/2022	31/07/2023	

Appendix B

B.1. ARTIC SARS-CoV-2 protocol (Tyson et al., 2020), currently V4

The NEBNext ARTIC SARS-CoV-2 FS Library Prep Kit for Illumina (E7650) contains the enzymes, buffers and oligonucleotides required to convert a broad range of total RNA into high quality, targeted, libraries for next-generation sequencing on the Illumina platform. Primers targeting the human EDF1 (NEBNext ARTIC Human Primer Mix 1) and NEDD8 (NEBNext ARTIC Human Primer Mix 2) genes are supplied as optional internal controls. The fast, user-friendly workflow also has minimal hands-on time. Each kit component must pass rigorous quality control standards, and for each new lot the entire set of reagents is functionally validated together by construction and sequencing of indexed libraries on an Illumina sequencing platform. For larger volume requirements, customized and bulk packaging is available by purchasing through the OEM/Bulks department at NEB.

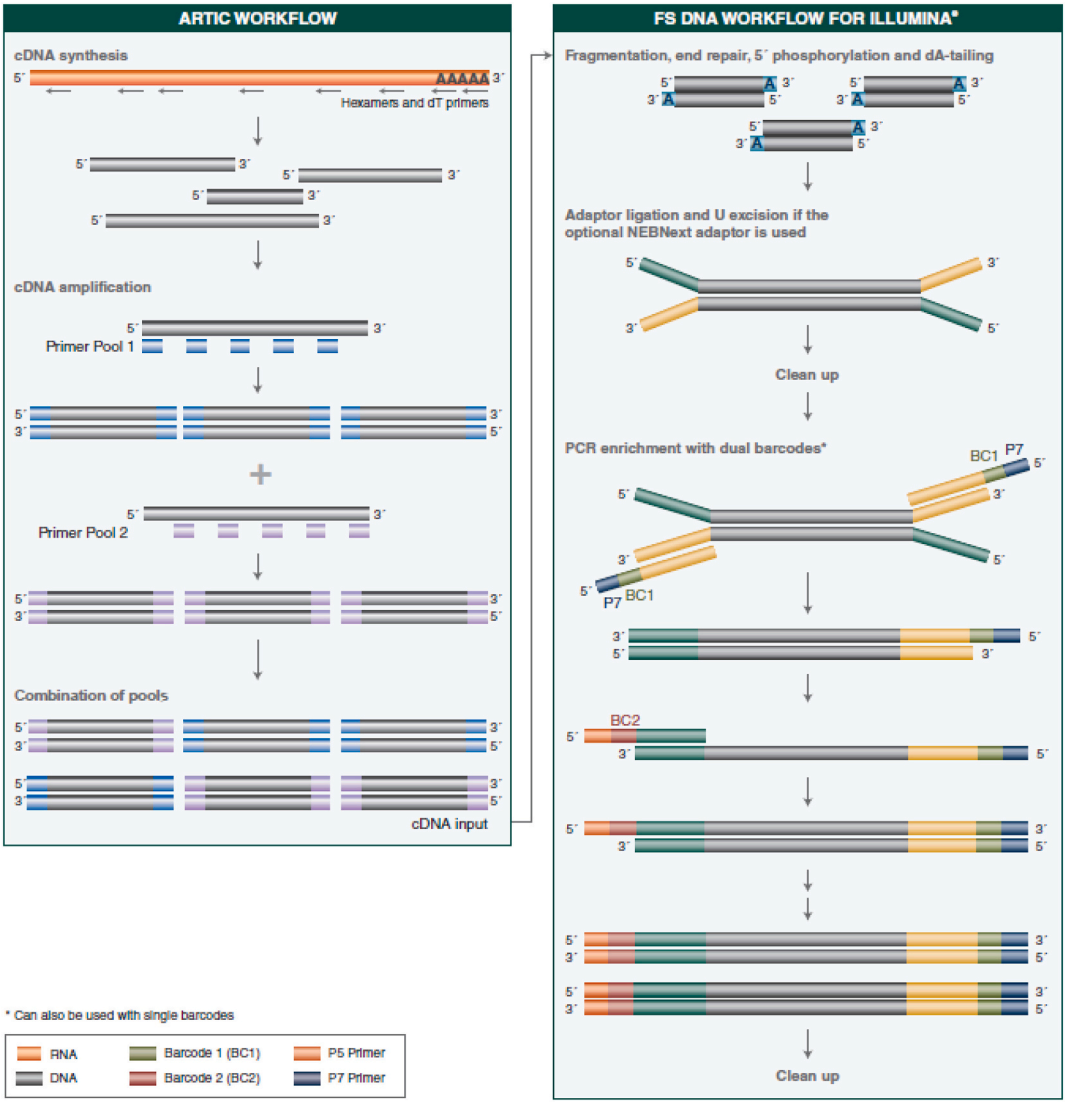


Fig. A.1. Workflow demonstrating the use of NEBNext ARTIC SARS-CoV-2 FS Library Prep Kit.

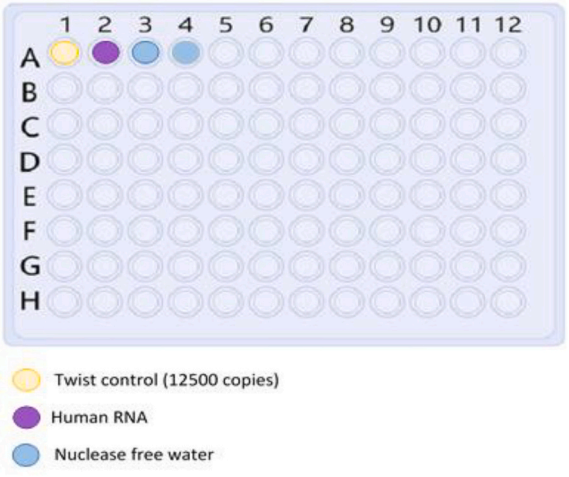
1. cDNA Synthesis

The presence of carry-over products can interfere with sequencing accuracy, particularly for low copy targets. Therefore, it is important to carry out the appropriate no template control (NTC) reactions to demonstrate that positive reactions are meaningful.

1.1. Gently mix and spin down the LunaScript RT SuperMix reagent. Prepare the cDNA synthesis reaction as described below:

	1 rxn	96-well plate
Component	Volume (µl)	Volume (µl)
RNA sample	8	–
LunaScript RT SuperMix	2	200
MM/strip tube:	25 µl	

- For no template controls, add water instead of RNA.
- Each RNA plate should have a no template (H_2O), a Human RNA sample and Twist RNA (12,500 copies) as controls.
- The wells to add the controls should be constant:



The controls should be switched each plate!

1.2. Seal the plate, vortex and spin prior to placing it in the thermocycler.

1.3. Incubate the reactions in a thermocycler with the following steps:

Cycle Step	Temperature	Time	Cycles
Primer Annealing	25 °C	2 min	1
cDNA Synthesis	55 °C	20 min	
Heat Inactivation	95 °C	1 min	
Hold	4 °C	∞	

*Set heated lid to 105 °C.
Samples can be stored at −20 °C for up to a week.

2. cDNA Amplification

Note: 4.5 µl cDNA input is recommended. If using < 4.5 µl of cDNA, add nuclease-free water to a final volume of 4.5 µl. We recommend setting up the cDNA synthesis and cDNA amplification reactions in different rooms to minimize cross-contamination of future reactions.

2.1. Gently mix and spin down reagents. Prepare the split pool cDNA amplification reactions as described below:

For Pool Set A:

	1 rxn	96-well plate
Component	Volume (µl)	Volume (µl)
cDNA	4.5	–
Q5 Hot Start High-Fidelity 2× MM	6.25	625
ARTIC SARS-CoV-2 Primer Mix 1	1.75	175
MM/strip tube:	100.0 µl	

For Pool Set B:

	1 rxn	96-well plate
Component	Volume (µl)	Volume (µl)
cDNA	4.5	–
Q5 Hot Start High-Fidelity 2× MM	6.25	625
ARTIC SARS-CoV-2 Primer Mix 2	1.75	175
MM/strip tube:	100.0 µl	

2.2. Seal the plate, vortex and spin prior to placing it in the thermocycler.

2.3. Incubate reactions in a thermocycler with the following steps:

Cycle Step	Temperature	Time	Cycles
Initial Denaturation	98 °C	30 s	1
Denaturation	95 °C	15 s	35
Annealing/Extension	63 °C	5 min	
Hold	4 °C	∞	1

*Set heated lid to 105 °C.

2.4. Combine the Pool A and Pool B PCR reactions for each sample.

2.5. Quality control

We used to perform Qubit as well as bioanalyzer, but we found that Qubit alone was predictive of whether libraries would succeed. We measured the controls and 8 random samples (2 samples from each source plate) with Qubit.

ID	Expected Qubit BR value
H_2O negative control	4–20
Human RNA negative control	4–20
Positive control	100–200
Sample	100–200

Samples can be stored at −20 °C for up to a week.

3. Fragmentation/End Prep

3.1. Mix 1.3 µl of the pooled amplicons (from step 2.4) with 11.7 µl of nuclease free water in a new 96well plate (Eppendorf TwinPlate) – Store the rest of the pools at −20 °C

3.2. Ensure that the Ultra II FS Reaction Buffer is completely thawed. If a precipitate is seen in the buffer, pipette up and down several times to break it up, and quickly vortex to mix. Place on ice until use.

3.3. Vortex the Ultra II FS Enzyme Mix 5–8 s prior to use and place on ice.

Note: It is important to vortex the enzyme mix prior to use for optimal performance.

3.4. Add the following components to a 0.2 ml thin wall PCR tube on ice:

Component	1 rxn Volume (µl)	96-well plate Volume (µl)
cDNA amplified	13	–
NEBNext Ultra II FS reaction Buffer	3.5	350
NEBNext Ultra II FS Enzyme Mix	1	100
MM/strip tube:	56.2 µl	

3.5. Seal the plate, vortex and spin prior to placing it in the thermocycler.

3.6. In a thermocycler, run the following program:

Temperature	Time
37 °C	25 min
65 °C	30 min
4 °C	∞

*Set heated lid to 75 °C.

If necessary, samples can be stored at −20 °C; however, a slight loss in yield (~20%) may be observed. We recommend continuing with adaptor ligation before stopping.

4. Adaptor Ligation
   4.1. Add the following components directly to the FS Reaction Mixture:

Component	1 rxn Volume (µl)	96-well plate Volume (µl)
FS reaction Mixture	17.5	–
NEBNext Adaptor for Illumina	1.25	125
NEBNext Ultra II Ligation MM	15	1500
MM/strip tube:	203.1 µl	

*Mix the Ultra II Ligation Master Mix by pipetting up and down several times prior to adding to the reaction.
**The NEBNext adaptor is provided in NEBNext Oligo kits.

4.2. Seal the plate, vortex and spin prior to placing it in the thermocycler
4.3. Incubate at 20 °C for 15 min in a thermocycler with the heated lid off.
4.4. Add 1.5 µl of • (red) USER® Enzyme to the ligation mixture from Step 4.3.
4.5. Mix well and incubate at 37 °C for 15 min with the heated lid set to ≥47 °C.

Samples can be stored overnight at −20 °C.

5. PCR Enrichment of Adaptor-ligated DNA/Dual Index PCR
   5.1. Add the following components to a sterile strip tube:

Component	1 rxn Volume (µl)	96-well plate Volume (µl)
Adaptor Ligated DNA Fragments	7.5	–
NEBNext Library PCR MM	12.5	1250
Dual Index Mix	5	–
MM/strip tube:	156.2 µl	

*Oligos from NEB dual index 96-well plate. 4 different sets.
**The leftover of the Adaptor Ligated DNA Fragments can be stored overnight @ –20 °C.

5.2. Seal the plate, vortex and spin before starting the program in the thermocycler.
5.3. Place the tube on a thermocycler and perform PCR amplification using the following PCR cycling Conditions:

Cycle Step	Temperature	Time	Cycles
Initial Denaturation	98 °C	30 s	1
Denaturation	98 °C	10 s	6*
Annealing/Extension	65 °C	75 s	
Final Extension	65 °C	5 min	1
Hold	4 °C	∞	

* Set heated lid to 105 °C. The number of PCR cycles recommended should be viewed as a starting point and may need to be optimized for particular sample types.

6. Pooling of libraries
   6.1. Without any QC step, all libraries from a plate should be pooled together.
   6.2. Pipette 4 µl of each library into a strip with a multichannel pipette and combine all volume into a 1.5 mL Eppendorf tube (~380 µl).

7. Cleanup of the pool
   7.1. Vortex NEBNext Sample Purification Beads to resuspend.
   7.2. Measure the pool volume and add 0.7× of resuspended beads to the PCR reaction. Mix well by pipetting up and down at least 10 times. Be careful to expel all of the liquid out of the tip during the last mix. Vortexing for 3–5 s on high can also be used. If centrifuging samples after mixing, be sure to stop the centrifugation before the beads start to settle out.
   7.3. Incubate samples on bench top for at least 5 min at room temperature.
   7.4. Place the tube/plate on an appropriate magnetic stand to separate the beads from the supernatant. If necessary, quickly spin the sample to collect the liquid from the sides of the tube or plate wells before placing on the magnetic stand.
   7.5. After 5 min (or when the solution is clear), carefully remove and discard the supernatant. Be careful not to disturb the beads that contain DNA targets.

   Note: do not discard the beads.

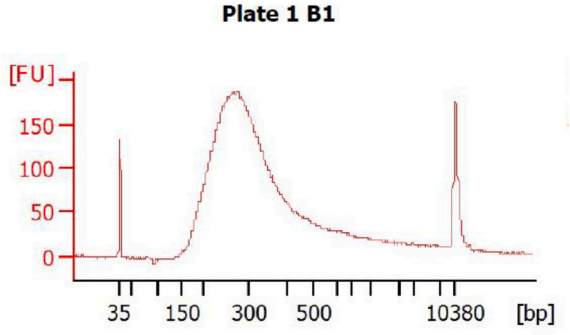
   7.6. Add 200 µl of 80% freshly prepared ethanol to the tube/plate while in the magnetic stand. Incubate at room temperature for 30 s, and then carefully remove and discard the supernatant. Be careful not to disturb the beads that contain DNA targets.
   7.7. Repeat Step 7.6. for a total of two washes. Be sure to remove all visible liquid after the second wash. If necessary, briefly spin the tube/plate, place back on the magnet and remove traces of ethanol with a p10 pipette tip.
   7.8. Air dry the beads for up to 5 min while the tube/plate is on the magnetic stand with the lid open.

   Note: Do not over-dry the beads. This may result in lower recovery of DNA.

   7.8.1. Remove the tube/plate from the magnetic stand. Elute the DNA target from the beads by adding 105 µl of 0.1× TE. Mix well by pipetting up and down 10 times, or on a vortex mixer. Incubate for at least 2 min at room temperature.
   7.8.2. If necessary, quickly spin the sample to collect the liquid from the sides of the tube or plate wells before placing back on the magnetic stand.
   7.8.3. Place the tube/plate on the magnetic stand. After 5 min (or when the solution is clear), transfer 100 µl to a new eppendorf tube.

   7.9. Repeat the cleanup steps with the 0.9× of resuspended beads.
   7.10. Elute the DNA target from the beads by adding 55 µl of 0.1× TE and mix well.
   7.11. Place the tube/plate on the magnetic stand. After 5 min (or when the solution is clear), transfer 50 µl to a new eppendorf tube and store at −20 °C.

8. Library QC
   8.1. Measure the concentration of the final pool with Qubit DNA BR kit.
   8.2. Assess the library size distribution with Agilent Bioanalyzer high sensitivity DNA reagents. The sample may need to be highly diluted.
   8.3. The expected concentration is between 100 and 200 ng/µl by BR Qubit.
   8.4. A peek sized of 200–250 bp is expected on a Bioanalyzer, based on a 30-min fragmentation time:



Example of final library size distributions on a Bioanalyzer.

References

Abdool Karim, S.S., de Oliveira, T., 2021. New SARS-CoV-2 Variants - Clinical, Public Health, and Vaccine Implications. N. Engl. J. Med. 384, 1866–1868.

Aksamentov, I., Roemer, C., Hodcroft, E.B., Neher, R.A., 2021. Nextclade: clade assignment, mutation calling and quality control for viral genomes. J. Open Source Software 6, 3773.

Baker, J.M., Nakayama, J.Y., O'Hegarty, M., McGowan, A., Teran, R.A., Bart, S.M., Mosack, K., Roberts, N., Campos, B., Paegle, A., McGee, J., Herrera, R., English, K., Barrios, C., Davis, A., Roloff, C., Sosa, L.E., Brockmeyer, J., Page, L., Bauer, A., Weiner, J.J., Khubbar, M., Bhattacharyya, S., Kirking, H.L., Tate, J.E., 2022. SARS-CoV-2 B.1.1.529 (omicron) variant transmission within households - four U.S. jurisdictions, November 2021-February 2022. MMWR Morb. Mortal Wkly. Rep. 71, 341–346.

Bosch, B.J., van der Zee, R., de Haan, C.A., Rottier, P.J., 2003. The coronavirus spike protein is a class I virus fusion protein: structural and functional characterization of the fusion core complex. J. Virol. 77, 8801–8811.

Center for Systems Science and Engineering (Johns Hopkins University), 2023. COVID-19 Dashboard [Online]. Available: https://coronavirus.jhu.edu/map.html [Accessed 19/01/2023].

Clarke, K.E.N., Jones, J.M., Deng, Y., Nycz, E., Lee, A., Iachan, R., Gundlapalli, A.V., Hall, A.J., MacNeil, A., 2022. Seroprevalence of infection-induced SARS-CoV-2 antibodies - United States, September 2021-February 2022. MMWR Morb. Mortal Wkly. Rep. 71, 606–608.

Ewels, P.A., Peltzer, A., Fillinger, S., Patel, H., Alneberg, J., Wilm, A., Garcia, M.U., Di Tommaso, P., Nahnsen, S., 2020. The nf-core framework for community-curated bioinformatics pipelines. Nat. Biotechnol. 38, 276–278.

Garrison, E., Marth, G., 2012. Haplotype-based variant detection from short-read sequencing. *arXiv*, p. 1207.

Grubaugh, N.D., Gangavarapu, K., Quick, J., Matteson, N.L., De Jesus, J.G., Main, B.J., Tan, A.L., Paul, L.M., Brackney, D.E., Grewal, S., Gurfield, N., Van Rompay, K.K.A., Isern, S., Michael, S.F., Coffey, L.L., Loman, N.J., Andersen, K.G., 2019. An amplicon-based sequencing framework for accurately measuring intrahost virus diversity using PrimalSeq and iVar. Genome Biol. 20, 8.

Hadfield, J., Megill, C., Bell, S.M., Huddleston, J., Potter, B., Callender, C., Sagulenko, P., Bedford, T., Neher, R.A., 2018. Nextstrain: real-time tracking of pathogen evolution. Bioinformatics 34, 4121–4123.

Havers, F.P., Reed, C., Lim, T., Montgomery, J.M., Klena, J.D., Hall, A.J., Fry, A.M., Cannon, D.L., Chiang, C.F., Gibbons, A., Krapiunaya, I., Morales-Betoulle, M., Roguski, K., Rasheed, M.A.U., Freeman, B., Lester, S., Mills, L., Carroll, D.S., Owen, S.M., Johnson, J.A., Semenova, V., Blackmore, C., Blog, D., Chai, S.J., Dunn, A., Hand, J., Jain, S., Lindquist, S., Lynfield, R., Pritchard, S., Sokol, T., Sosa, L., Turabelidze, G., Watkins, S.M., Wiesman, J., Williams, R.W., Yendell, S., Schiffer, J., Thornburg, N.J., 2020. Seroprevalence of Antibodies to SARS-CoV-2 in 10 Sites in the United States, March 23-May 12, 2020. JAMA Intern. Med. 180, 1576–1586.

Jørgensen, S.B., Nygård, K., Kacelnik, O., Telle, K., 2022. Secondary attack rates for omicron and Delta variants of SARS-CoV-2 in Norwegian households. Jama 327, 1610–1611.

Li, H., 2011. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. Bioinformatics 27, 2987–2993.

Li, H., Durbin, R., 2009. Fast and accurate short read alignment with burrows-wheeler transform. Bioinformatics 25, 1754–1760.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., 2009. The sequence alignment/map format and SAMtools. Bioinformatics 25, 2078–2079.

Mallm, J.-P., Bundschuh, C., Kim, H., Weidner, N., Steiger, S., Lander, I., Börner, K., Bauer, K., Hübschmann, D., Benes, V., Rausch, T., de Azevedo, N.T.D., Telzerow, A., Jost, K.L., Parthé, S., Schnitzler, P., Boutros, M., Müller, B., Bartenschlager, R., Kräusslich, H.-G., Rippe, K., 2021. Local emergence and decline of a SARS-CoV-2 variant with mutations L452R and N501Y in the spike protein. medRxiv 2021, 04.27.21254849.

McLaren, W., Gil, L., Hunt, S.E., Riat, H.S., Ritchie, G.R., Thormann, A., Flicek, P., Cunningham, F., 2016. The Ensembl variant effect predictor. Genome Biol. 17, 122.

Ministerium für Soziales, G. u. I. B.-W, 2021. Mutiertes Coronavirus: Erster Nachweis der Südafrika-Variante B.1.351 in Baden-Württemberg.

Patel, H.V.S., Monzón, S., Espinosa-Carrasco, J., Heuer, M.L., Gabernet, G., Julia, M., Kelly, S., Sameith, K., Garcia, M., 2020. nf-core/viralrecon: nf-core/viralrecon v1.1.0 - Steel Pangolin.

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. Nat. Microbiol. 5, 1403–1407.

Rausch, T., Hsi-Yang Fritz, M., Korbel, J.O., Benes, V., 2019. Alfred: interactive multi-sample BAM alignment statistics, feature counting and feature annotation for long- and short-read sequencing. Bioinformatics 35, 2489–2491.

Rausch, T., Fritz, M.H., Untergasser, A., Benes, V., 2020. Tracy: basecalling, alignment, assembly and deconvolution of sanger chromatogram trace files. BMC Genomics 21, 230.

Robert Koch Institute (RKI), 2021a. Deutscher Elektronischer Sequenzdaten-Hub (DESH) [Online]. Available: https://www.rki.de/DE/Content/InfAZ/N/Neuartiges_Coronavirus/DESH/DESH.html [Accessed 06/12/2021].

Robert Koch Institute (RKI), 2021b. *Qualitätsvorgaben für die Sequenzdaten* [Online]. Available: https://www.rki.de/DE/Content/InfAZ/N/Neuartiges_Coronavirus/DESH/Qualitaetskriterien.pdf [Accessed 06/12/2021].

Robert Koch Institute (RKI), 2021c. SARS-CoV-2 Varianten in Deutschland [Online]. Available: https://public.data.rki.de/t/public/views/IGS_Dashboard/DashboardSublineages?%3Aembed=y&%3AisGuestRedirectFromVizportal=y [Accessed 31/01/2024].

Shu, Y., McCauley, J., 2017. GISAID: global initiative on sharing all influenza data - from vision to reality. Euro Surveill. 22.

Stringhini, S., Wisniak, A., Piumatti, G., Azman, A.S., Lauer, S.A., Baysson, H., De Ridder, D., Petrovic, D., Schrempft, S., Marcus, K., Yerly, S., Arm Vernez, I., Keiser, O., Hurst, S., Posfay-Barbe, K.M., Trono, D., Pittet, D., Gétaz, L., Chappuis, F., Eckerle, I., Vuilleumier, N., Meyer, B., Flahault, A., Kaiser, L., Guessous, I., 2020. Seroprevalence of anti-SARS-CoV-2 IgG antibodies in Geneva, Switzerland (SEROCoV-POP): a population-based study. Lancet 396, 313–319.

Tyson, J.R., James, P., Stoddart, D., Sparks, N., Wickenhagen, A., Hall, G., Choi, J.H., Lapointe, H., Kamelian, K., Smith, A.D., Prystajecky, N., Goodfellow, I., Wilson, S.J., Harrigan, R., Snutch, T.P., Loman, N.J., Quick, J., 2020. Improvements to the ARTIC multiplex PCR method for SARS-CoV-2 genome sequencing using nanopore. *bioRxiv*.

Urhan, A., Abeel, T., 2021. Emergence of novel SARS-CoV-2 variants in the Netherlands. Sci. Rep. 11, 6625.

Wood, D.E., Lu, J., Langmead, B., 2019. Improved metagenomic analysis with kraken 2. Genome Biol. 20, 257.

World Health Organization (WHO), 2020a. Director-General's remarks at the media briefing on 2019-nCoV on 11 February 2020.

World Health Organization (WHO), 2020b. Tracking SARS-CoV-2 variants [Online]. Available: https://www.who.int/en/activities/tracking-SARS-CoV-2-variants/ [Accessed 31/01/2024].

World Health Organization (WHO), 2023. Coronavirus disease (COVID-19) Weekly Epidemiological Update and Weekly Operational Update.

Xia, S., Zhu, Y., Liu, M., Lan, Q., Xu, W., Wu, Y., Ying, T., Liu, S., Shi, Z., Jiang, S., Lu, L., 2020. Fusion mechanism of 2019-nCoV and fusion inhibitors targeting HR1 domain in spike protein. Cell. Mol. Immunol. 17, 765–767.

Yue, C., Song, W., Wang, L., Jian, F., Chen, X., Gao, F., Shen, Z., Wang, Y., Wang, X., Cao, Y., 2023. Enhanced Transmissibility of XBB.1.5 is Contributed by Both Strong ACE2 Binding and Antibody Evasion. *bioRxiv*, 2023.01.03.522427.