Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Scalable regional genomic surveillance reveals previously unrecognized SARS-CoV-2 recombinant diversity and limited onward transmission

Christian Kohler

christian.kohler@med.uni-greifswald.de

University Medicine Greifswald https://orcid.org/0000-0003-3921-6776

Ana Tzvetkova

University Medicine Greifswald

Anne Pohlmann

Friedrich-Loeffler-Institute https://orcid.org/0000-0002-5318-665X

Kathrin Lehmann

University Medicine Greifswald

Tilo Sasse

LAGuS – Landesamt für Gesundheit und Soziales Mecklenburg-Vorpommern

Veronika Balau

IMD Labor Greifswald MVZ Labor Greifswald GmbH

Martin Beer

Friedrich-Loeffler-Institute https://orcid.org/0000-0002-0598-5254

Lars Kaderali

University Medicine Greifswald https://orcid.org/0000-0002-2359-2294

Katja Goller

University Medicine Greifswald

Nils-Olaf Hübner

University Medicine Greifswald

Karsten Becker

Greifswald University Medicine https://orcid.org/0000-0002-6391-1341

Biological Sciences - Article

Keywords: SARS-CoV-2, recombinants, omicron, whole genome sequencing, surveillance, genomic epidemiology, Mecklenburg-Western Pomerania

Posted Date: August 10th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10510603/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: There is **NO** Competing Interest.

Abstract

Global SARS-CoV-2 surveillance based on consensus genome sequences has enabled unprecedented tracking of viral spread and evolution but provides limited resolution of intra-sample diversity and limited sensitivity for the reliable detection of recombinant genomes. Here, we analyzed 3,360 SARS-CoV-2 genomes generated through regional surveillance in northeastern Germany (Mecklenburg–Western Pomerania) between 2022 and 2025. Reconstruction of local spatio-temporal variant dynamics revealed substantial lineage diversity, regional deviations from global circulation patterns, and periods of intense co-circulation favoring co-infection and recombination. To systematically identify recombinant genomes, we implemented a two-stage framework combining high-sensitivity consensus-based screening (REcombination BARcode detector REBAR) with targeted raw-read validation. Among 61 candidate recombinants, 21 were supported by read-level evidence, including 17 previously unrecognized genomes. Recombinant breakpoints clustered in functionally relevant genomic regions, particularly around ORF1b and the spike gene, indicating non-random patterns of genome exchange. Despite frequent co-infections, only a small subset of recombinants showed evidence of onward transmission, suggesting that successful establishment represents a major constraint on recombinant SARS-CoV-2 evolution. Our findings demonstrate that recombination-aware genomic surveillance complements routine consensus genome surveillance by resolving intra-sample diversity and recombinant diversity while remaining scalable for population-level implementation.

Introduction

Despite ongoing global efforts toward pandemic preparedness, the emergence of Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) marked a strong reminder of the world's continued vulnerability to novel viral threats. First identified in Wuhan, China, in December 2019, this novel betacoronavirus caused an unusual viral pneumonia later termed as Coronavirus Disease 2019 (COVID-19) [1]. To date (February 2026), the pandemic has spread globally, with more than 778 million confirmed COVID-19 infections and over 7.1 million officially reported deaths worldwide (https://data.who.int/dashboards/covid19). Driven by the urgent need to understand the virus, cutting-edge genomic analysis techniques and protocols rapidly emerged, proving crucial for tracing the origin, transmission, and evolution of SARS-CoV-2, identifying emerging variants, and guiding public health responses [2], thereby enabling accurate real-time surveillance of the pandemic.

Initially, closely related early SARS-CoV-2 lineages originating from the Wuhan strain predominated; however, through genetic drift and selective pressures [3], the virus rapidly accumulated mutations, resulting in the successive emergence of more transmissible variants that drove multiple waves of infection. In particular, the first Variants of Concern (VOCs), including Alpha (B.1.1.7), Beta (B.1.351), Gamma (P.1), and Delta (B.1.617.2), were associated with increased transmissibility and, in some cases, greater disease severity, contributing to elevated mortality rates [4–7]. Together with additional variants, these early VOCs shaped the first two years of the pandemic through December 2021, during which approximately 289 million confirmed cases and more than 5.4 million confirmed deaths were reported

globally (https://data.who.int/dashboards/covid19). From late 2021 into early 2022, the Omicron variant and its sublineages rapidly displaced all preceding VOCs and have remained dominant worldwide. Compared with earlier VOCs, Omicron exhibits markedly increased transmissibility but reduced intrinsic virulence, resulting in a lower proportion of severe disease and mortality per infection [8–11]. Between early January 2022 and early 2026, an additional ~ 490–500 million confirmed infections and approximately 1.6 million confirmed deaths were recorded globally, with the majority of new cases and deaths occurring between early 2022 and early 2023, after which numbers sharply declined (https://data.who.int/dashboards/covid19). The comparatively smaller rise in mortality despite the substantial increase in case numbers reflects not only the reduced intrinsic virulence of Omicron but also the protective effects of widespread vaccination and population-level immunity acquired through prior SARS-CoV-2 infections [12,13]. Interestingly, during the Omicron wave, a marked increase in the emergence of recombinant SARS-CoV-2 lineages was observed, being initially most detected in North America and Northern Europe, particularly in the UK, before subsequently spreading to other regions worldwide [3]. Recombination, alongside reassortment, is a key evolutionary mechanism of RNA viruses and occurs in both segmented and non-segmented viruses [14–16] and requires co-circulation and co-infection of the same host by different viral lineages. During the transition from Delta to Omicron, and particularly throughout the subsequent Omicron waves, the unprecedented number of infections together with the extensive co-circulation of genetically diverse Omicron sublineages substantially increased the likelihood of co-infections, thereby creating favorable conditions for SARS-CoV-2 recombination. The combination of high transmission dynamics and substantial viral diversity likely contributed to the marked rise in recombinant lineages observed during this period [3]. Some descendants of these recombinant viruses have shaped, and continue to shape the course of SARS-CoV-2 infections worldwide (https://data.who.int/dashboards/covid19/variants).

Unfortunately, as emergency funding and interest waned following the acute phase of the COVID 19 pandemic, routine SARS-CoV-2 genomic surveillance has been scaled back or integrated into broader respiratory virus monitoring efforts. Moreover, widespread diagnostic testing, which supplied many of the samples for sequencing, has declined, resulting in more sporadic and incomplete tracking of ongoing SARS CoV 2 evolution [17]. Therefore, it is even more important to establish and maintain not only global but also regional and local surveillance systems to monitor SARS-CoV-2 and, in the future, further emerging infectious pathogens. The ongoing project “CoMV-Gen,” launched in January 2021 [18], continues to monitor SARS-CoV-2 in the federal state of Mecklenburg-Western Pomerania (M-V), Germany, a region comprising mainly rural areas in northeastern Germany. Here, we analyzed full-genome sequencing data and associated metadata collected between January 2022 and January 2025. Using the REcombination BARcode detector (REBAR, https://github.com/phac-nml/rebar), we tracked SARS-CoV-2 variant dynamics in M-V while systematically screening for recombinant lineages. Targeted raw-read analyses distinguished recombinants from co-infections, establishing a scalable two-stage strategy for large-scale genomic surveillance. This approach revealed substantial recombinant diversity, while showing that only a subset of recombinants achieved detectable onward transmission. We show that a substantial component of ongoing SARS-CoV-2 evolution remains hidden from conventional

genomic surveillance and present a practical framework to uncover this overlooked layer of viral diversity.

Results

Sample Composition and Metadata Summary

Between January 1, 2022, and January 11, 2025, 4,675 SARS-CoV-2–positive patient samples were sequenced. Of these, 3,360 (72.8%) achieved ≥ 99% genome coverage with a minimum depth of > 20-fold and were included in downstream analyses. Most samples (64.4%, n = 2,166) were obtained from University Medicine Greifswald (UMG). The remaining samples were provided by three sites in M-V: LAGuS Rostock (21.4%, n = 719), IMD Greifswald (12.9%, n = 433) and Dietrich-Bonhoeffer Hospital Neubrandenburg (1.3%, n = 44). LAGuS collected samples from private practices and major hospitals across the state. In total, 2,930 samples (87.2%) were derived from residents of the study region M–V, whereas 212 samples originated from individuals whose primary residence was outside the state but who were tested within M–V. Among these non-resident cases, most were linked to the neighboring federal state of Brandenburg (n = 96; 45.3%). The remaining 116 individuals resided in other German federal states (Table S1). Six samples were obtained from patients registered in Poland (n = 6) and one from the Netherlands. For 211 samples, the patients' place of residence was unknown; these samples were nevertheless included in the analysis because they were isolated in M–V.

Of the 2,930 samples originating from M–V, most were from the districts Vorpommern-Greifswald (n = 2,142, 73.1%), followed by Vorpommern-Rügen (n = 464, 15.8%) and Mecklenburgische Seenplatte (n = 188, 6.4%). The remaining 136 samples (4.6%) were assigned to other districts (Fig. 1a). This coverage indicates that the eastern part of M–V was well represented, whereas the western region was sparsely sampled, largely due to the discontinuation of mandatory SARS-CoV-2 sample submission from early-middle 2023 onward.

Spatio-temporal distribution and high diversity of SARS-CoV-2 clades and lineages in M-V

To classify individual sequences, Nextclade [19] was used, and associated sampling data were employed to reconstruct the temporal succession of viral clades from 2022 to 2025 (Fig. 1b). The temporal patterns of sequence count closely followed the reported trends in COVID-19 incidence in M-V (Fig. 1b, Figure S1, table S2 and https://corona-zahlen-heute.de/deutschland/mecklenburg-vorpommern/), suggesting that our sequencing captured the dynamics of viral circulation over time rather than absolute case numbers. During the study period, 29 distinct Nextstrain clades were identified. The most prevalent were clade 22B (BA.5; n = 1,070, 31.9%), Clade 21K (BA.1; n = 453, 13.5%), and clade 21L (BA.2; n = 389, 11.6%), reflecting the temporal sampling. Other clades were less frequent, including Clade 23A (7.0%), while Clade 21A (Delta) was detected in only a single sample (Table S3). At the onset of the study period,

circulating Nextstrain clades declined, consistent with the rapid displacement of Delta by emerging Omicron variants (Fig. 1b). From May 2022, genetic diversity expanded substantially, reflecting ongoing clade diversification. A brief decline occurred in September 2022, when only clade 22B was detected, after which clade diversity rose again and remained elevated through July 2023. Diversity peaked between October and December 2023 before declining, likely reflecting both decreasing case numbers and reduced sample availability. The final wave between June and December 2024 was again characterized by concurrent circulation of multiple distinct clades (Fig. 1b, Table S3).

In January 2022, 35.6% of detected sequences belonged to Delta lineages, dropping to 3.5% in February (Fig. 1b, table S3). From then on, only Omicron variants were observed. Clade 21K (BA.1) dominated from January to March 2022, followed by 21L (BA.2) until June, and 22B (BA.5) from July 2022, remaining dominant through February 2023. Variants corresponding to clades 22D and 22E were detected from October 2022 onward, but at low frequencies. From January 2023 until November 2023, SARS-CoV-2 circulation was strongly influenced by XBB-recombinant lineages (clades 22F, 23A, 23B, 23D, 23E, and 23G). Thereafter, circulation was largely driven by non-recombinant clades such as 23F, 23H, 23I, and 24A, 24B, 24C, and 24E and 24G. The newly emerged recombinant clade XEC was first detected in August 2024 and continued to be identified through January 2025 (Fig. 1b, Table S3). This regional timeline largely mirrors the global pandemic pattern, although some deviations from global and European trends were observed (Fig. 1c). The proportion of Delta variants (clades 21A, 21I, and especially 21J) in January–February 2022 exceeded the global average, and BA.1 variants (clade 21K) remained more prevalent through March–April 2022 compared with Europe and worldwide. Thereafter, the spread and timing of clade emergence in M–V followed broader trends, although some clades, including 24D, 24I, 24H, and 25H, were not detected locally (Fig. 1c).

Finally, sequences were assigned to individual lineages using the Pangolin tool [20,21]. Across the 29 clades, 423 Pangolin lineages were identified. The most common clades/lineages were BA.1.1 (n = 258, 7.7%), BA.2 (n = 253, 7.5%), BA.5.2 (n = 155, 4.6%), BF.7 (n = 139, 4.1%), BA.5.1 (n = 136, 4.1%), BE.1.1 (n = 110, 3.3%), and BA.5.2.1 (n = 105, 3.1%). Notably, lineages detected only once (n = 170, 40.2%) or twice (n = 73, 17.3%) outnumbered those represented by three or more isolates (Table S3 and S4). These findings indicate substantial lineage diversity within the clades, which was further supported by phylogenetic analysis (Fig. 1d). Phylogenetic analyses indicated a largely homogeneous population structure, with no substantial fragmentation of established clades and no discernible clustering according to geographic origin, including M-V and other German states as well as the wider European dataset.

Dynamics and Detection of SARS-CoV-2 Recombinants in M–V

SARS-CoV-2 circulation in M–V shifted from the phylogenetically distinct Delta clade to Omicron in January–February 2022, followed by continued diversification within Omicron lineages through January 2025. Initial lineage assignment using Nextclade and Pangolin nomenclature (Reference: Wuhan-Hu-

1/2019 (MN908947); Updated at: 2025-03-26) identified 541 recombinant sequences (16.1%) over the study period (Table S4). High-prevalence recombinants, including XBB and XEC lineages, were assigned to distinct Nextstrain clades, whereas 76 lower-prevalence sequences were grouped into a placeholder “recombinant” clade comprising 21 Pangolin lineages (Table S5).

Given the ongoing evolution of nomenclature systems and the potential underrepresentation of lineage-defining genomic positions, emerging, misclassified, or previously unrecognized recombinant lineages may not be fully captured by assignment based solely on Nextclade and Pangolin. To address this further and in more detail, all 3,360 sequences were screened for recombination events using the REBAR algorithm, which detects recombination based on lineage-defining mutation barcodes derived from consensus sequences. Additional raw-read–level analysis was performed to distinguish true recombinant genomes from potential co-infections, contamination, or consensus-calling artefacts (Fig. 2a and 2b). In total, 815 sequences were initially identified as recombinant candidates (Table S6). Of these, 348 sequences harbored additional mutations likely acquired after the initial recombination events and were classified as XBB-derived descendants rather than direct recombinants, and were excluded from further analysis.

A total of 406 sequences were assigned to previously described recombinant lineages, encompassing 74 distinct lineages (Table S7). Predominant lineages included XBB-derived lineages (n = 189), XEC (n = 142), XCT (n = 23), XEK (n = 19), and XBL (n = 9), while the remainder occurred at low frequency or as single instances (Table S7). REBAR also identified six sequences previously classified as non-recombinant by Nextstrain/Pangolin, including XBB.1.5.12 (20233305431), XEC.21 (20240205825), XDS (20234905736), XBA (20220101659), XD (20220101664), and XEC (20240606217), and reassigned two sequences (2022072286 and 20231105078) to XY and XCA, instead of XE and XBS, respectively (Table S6).

In addition, 53 previously unannotated recombinant candidates were detected. All 61 candidate sequences were subjected to additional variant analysis at lineage-defining genomic positions using raw-read data, alongside comparative controls representing previously described recombinant lineages (Table S8). This analysis revealed 25 putative true recombinants, of which four (20231105078, 20234905736, 20240205825, 20233305431) could be assigned to previously described lineages (XCA, XDS, XEC.21, and XBB.1.5.12). The remaining 21 sequences represent 17 novel, hitherto uncharacterized recombinants, highlighting ongoing genomic diversification and the emergence of previously unrecognized evolutionary trajectories (Table S8, Figure S2). This represents 0.625% of all analyzed sequences.

From the beginning of 2022 until the end of April, four novel recombinants (20220301923, 20220201848, 20220702180, 20220702198) were identified, arising from parental lineages within Delta and Omicron. Three additional Omicron recombinants, derived from BA.1 and BA.2 parental lineages (20220101711, 20220702224, 20221903176), were also detected (Table S8, Fig. 2c). From May to December 2022, newly identified putative recombinants were exclusively associated with BA.5 lineage parents (Table S8,

Fig. 2c). With the exception of a single novel recombinant detected in December 2023 (20240205818), no further novel recombinants were observed; all remaining sequences corresponded to previously described or internationally reported recombinant lineages (Table S8, Fig. 2c). Notably, recombinant sequences increased in frequency throughout 2023, with XBB-derived lineages predominating until November 2023, followed by a transient rise in XCT lineages. Non-recombinant lineages (e.g., JN.1, BA.2.86, HK.3) subsequently regained dominance (Table S3, Fig. 1b). Following a three-month period of low case numbers (March–May 2024), recombinants derived from XEC and XEK lineages accounted for a substantial proportion of sequences from August 2024 onwards and remained detectable until January 2025 (Fig. 2c).

Recombinant breakpoint distribution and hotspot analysis

A total of 21 putative novel recombinant genomes of SARS-CoV-2 and 401 circulating recombinant lineages (CRLs) were analyzed based on breakpoint predictions generated by REBAR (Fig. 3). Across all samples, 70 partly overlapping but distinct breakpoints intervals (DBPs) were identified, comprising 27 DBPs in the novel recombinant group and 43 DBPs among CRLs (Table S9). Within the novel recombinant group, 8 sequences (38%) harbored a single DBP, whereas 13 sequences (62%) exhibited two DBPs (Fig. 3a, Table S8). By contrast, the majority of CRLs (92.5%, n = 371) contained only a single breakpoint. Only a small fraction displayed multiple breakpoints, including two breakpoints in 6.7% (n = 27), three breakpoints in 0.5% (n = 2), and four breakpoints in 0.25% (n = 1) of cases, respectively (Table S11).

Within the novel recombinant lineage group, breakpoint positions were not randomly distributed across the genome but instead exhibited clear clustering within specific regions (Fig. 3a). The most prominent hotspot was located within the spike (S) gene, between nucleotide positions 21,764 and 22,103 (Fig. 3b), where the highest density of breakpoints was observed. Recurrent breakpoint intervals, including regions spanning approximately ~ 22,600–27,437, were detected across multiple independent samples, further supporting this region as a recombination hotspot. A second major hotspot was identified at the transition between ORF1b and the spike gene (~ 18,500–22,000). In addition, a third, less frequently reported region of breakpoint clustering was detected within the early ORF1a region (~ 2,000–7,000), where multiple independent recombination events were observed, including recurrent intervals around 2,300–6,500 (Fig. 3b). Statistical evaluation using a permutation-based framework demonstrated that the observed clustering of breakpoints deviates markedly from a random genomic distribution. Significant enrichment of breakpoints was detected at the ORF1b–spike transition (genomic regions 19,631–20,150 and 20,741–21,555) and within the spike gene (21,563–22,990 and 24,410–25,090) ($p < 0.05$; Table S10). In contrast, the early ORF1a region did not show statistically significant enrichment ($p > 0.05$), despite recurrent observations across three independent samples, suggesting a potential but less robust hotspot. Because statistical testing was performed across the entire genome (29,903 positions), correction for multiple comparisons was required to control for false positives. We therefore applied the Benjamini–Hochberg procedure to control the false discovery rate (FDR). After adjustment, statistically

significant signals (FDR < 0.05) were confined to small regions within the spike gene (Fig. 3b and Table S10). The restriction of significance to these narrow intervals is likely attributable to the limited sample size combined with variability in breakpoint positions.

Comparable breakpoint hotspot patterns were also observed among circulating recombinant lineages (CRLs), including the frequently detected XBB, XEC and XCT lineages, as well as less prevalent CRLs (Fig. 3b, Table S9, Table S10). The most common CRLs exhibited breakpoint enrichment at the ORF1b–spike junction, within the spike gene, and at the ORF3a–ORF7b junction (Fig. 3b, Table S10). In addition, significant enrichment at the ORF1a–ORF1b junction was observed in CRLs but was not detected in the dataset analyzed here (Fig. 3b, Table S10). Genomes containing two or more breakpoints were observed and tended to comprise smaller exchanged genomic segments, whereas genomes with a single breakpoint consisted of two larger regions derived from the respective parental lineages (Figure S2).

In addition to the analysis of the overall distribution of breakpoint intervals across the genome, each annotated gene was tested separately to determine whether the proportions of breakpoint intervals overlapping the corresponding gene differ between NRLs and CRLs (Table 1). Fisher's exact test was applied, and p-values were corrected for testing across eight genes using the Benjamini-Hochberg procedure.

The only gene showing a significant difference was the S gene, with 59.01% of breakpoint intervals in CRLs overlapping this gene compared to 31.94% in NRLs (FDR 0.00012). The ORF1b gene showed a nominally significant p-value (0.028), with 19.44% of breakpoint intervals in NRLs overlapping this gene compared to 10.25% in CRLs; however, the FDR exceeded the significance level. A comparison of breakpoint distributions between NRLs and CRLs revealed a striking shift in genomic localization. In contrast, established circulating recombinant lineages exhibited a pronounced enrichment of breakpoints within the spike gene. Conversely, NRLs were distinguished by a considerably higher proportion of breakpoint intervals in ORF1ab. However, given the limited number of NRLs, this interpretation remains speculative and requires validation in larger datasets. In addition to the small sample size of NRLs, another drawback/limitation of this method should be considered: breakpoint intervals overlapping more than one gene were counted for each overlapping gene, as the exact breakpoint position could not be determined precisely.

Evidence for co-infections from REBAR and raw read analysis

Based on REBAR and raw read analyses of the 61 selected sequences, 36 sequences were classified as potential co-infections. These occurred predominantly during the transition phase from Delta to Omicron between January and March 2022, accounting for approximately 70% (n = 25) of cases. Among these, 22 sequences showed evidence of the co-existence of Delta and Omicron variants within the same sample, while three samples likely represented co-infections involving two distinct Omicron lineages (Figure S3). Additional putative Delta–Omicron co-infections were detected only sporadically thereafter, with four cases identified in November–December 2022 and March 2023. The remaining six samples were

consistent with mixed infections involving different Omicron lineages. Table S11 summarizes all sample metadata and analysis results.

Discussion

The COVID-19 pandemic marked an unprecedented expansion of global genomic surveillance, enabling near real-time tracking of SARS-CoV-2 evolution through standardized frameworks such as Nextstrain [19] and PANGOLIN [20], supported by large-scale data sharing via repositories including GISAID [22], GenBank [23] and others. While highly effective at scale, these approaches rely predominantly on consensus genome sequences, thereby limiting the resolution of intra-sample variation and the detection of complex evolutionary processes such as recombination [24,25]. Although raw sequencing data are available through repositories such as the European Nucleotide Archive [26], their integration into routine surveillance remains limited [24,25], with substantially fewer datasets deposited as FASTQ compared to consensus FASTA genomes. In addition, global surveillance is constrained by uneven sampling and heterogeneous data generation, which may obscure fine-scale evolutionary dynamics [27,28]. In this context, consensus genomes remain the most widely available and standardized format, enabling scalable analyses and cross-study comparability despite known resolution limits. Accordingly, we analyzed 3,360 complete SARS-CoV-2 genome sequences from a geographically confined population, integrating genomic, phylogenetic and spatio-temporal approaches, including lineage assignment using PANGOLIN, clade tracking with Nextstrain, and phylogenetic reconstruction. In addition, we applied the consensus-based recombination detection algorithm REBAR as a sensitive first-pass screening tool, enabling prioritization of candidate recombinant genomes for downstream validation. Leveraging its high sensitivity for comprehensive candidate detection [29–31], REBAR identified a restricted subset of candidate recombinant genomes, which enabled efficient downstream raw-read-level analyses by focusing validation exclusively on these selected candidates rather than the full dataset.

Initial analyses revealed region-specific patterns that differed from contemporaneous European and global SARS-CoV-2 trends, potentially reflecting local transmission dynamics and heterogeneous sampling (Fig. 1). Such differences may have been shaped by temporally and regionally variable non-pharmaceutical interventions (NPIs), ranging from strict mobility restrictions to less stringent, recommendation-based approaches [32–34], as exemplified by cross-country differences including Sweden [35]. More recent analyses further highlight that NPI effectiveness is highly context-dependent, influenced by behavioral responses and implementation strategies [35,36]. In Germany, including M-V, NPIs were implemented in a prolonged and intermittently stringent manner, with most legally binding measures phased out by early to mid-2022 and residual provisions expiring in April 2023. Following relaxation, mobility likely increased but remained below pre-pandemic levels, consistent with sustained behavioral adaptation. In parallel, vaccination strategies, coverage heterogeneity and waning immunity may have further shaped lineage dynamics [37–39]. More recent years also saw reduced and uneven sequencing intensity, particularly after 2023, further limiting surveillance resolution [40–43], with persistent disparities

in sequencing capacity across regions [40,44]. Collectively, these factors underscore the value of high-resolution regional genomic surveillance, which enables fine-scale reconstruction of transmission dynamics and provides a critical complement to global datasets.

Our high-resolution dataset provided an opportunity to investigate evolutionary processes that are often underrepresented in large-scale surveillance, particularly recombination [45–48]. As most laboratories generate consensus genomes, we used a consensus-based screening approach for initial candidate identification. Although multiple recombination frameworks exist, comprehensive benchmarking across methods is rarely feasible in large-scale surveillance and was beyond the scope of this study [29,30,47]. REBAR was therefore selected as a sensitive screening tool, identifying 61 candidate recombinant genomes (1.8%). This proportion exceeds estimates from strictly validated datasets (~ 0.2–0.5%) [25,29] but is lower than broader screening-based estimates (~ 2–3%) [47], likely reflecting both high sensitivity and susceptibility to artefacts such as mixed infections or amplification bias. All candidates were subsequently validated using raw-read data to distinguish true recombination from alternative explanations.

Through this approach, 21 sequences (0.6% of all sequences) were assigned to putative novel recombinant genomes, including 17 previously undescribed recombinants (Figure S4). During the transition from Delta to Omicron, we detected 22 co-infections, reflecting substantial co-circulation of both variants and providing the prerequisite for recombination, consistent with previous reports [25]. Despite these favorable conditions, only four Delta–Omicron recombinants were identified (Fig. 2C, Fig. 3, Figure S2, Figure S4), underscoring the limited emergence of inter-variant recombinants. In these cases, recombination yielded genomes with spike proteins largely derived from the Delta lineage (Figure S2). The restricted spread of these recombinants likely reflects a combination of immunological and epidemiological constraints. As the spike protein constitutes the principal target of neutralizing antibodies against SARS-CoV-2, with most neutralizing activity directed against epitopes within this protein [49–51], population-level immunity in M-V, which has been shaped by the predominance of Delta variants between July and December 2021 [18], may have limited any selective advantage of recombinants harboring Delta-derived spike proteins. This constraint would have been further reinforced by the emergence of Omicron, which exhibited enhanced immune escape. In addition, recombination within ORF1ab, observed in two of these Delta–Omicron recombinants, may have influenced viral fitness, given the role of this region in encoding the replication–transcription machinery [52,53]. However, the functional consequences of such exchanges remain difficult to predict and are likely contingent on genomic compatibility [54], and therefore require experimental validation.

All other novel recombinant genomes identified here involved Omicron-derived parental lineages. Recombination within the Omicron lineage was more frequently observed, consistent with increased genomic compatibility among closely related variants, as reported elsewhere [55]. Notably, despite identifying only 11 Omicron–Omicron co-infections, we detected 13 corresponding recombinant lineages (Figure S3), suggesting underdetection of co-infections, as our analysis prioritized recombinant

identification using REBAR rather than systematic co-infection screening. Two NRLs were each detected independently on three occasions (2022804576, 2022804577, 2022804637 and 20223203839, 20223203861, 20223203862), indicative of limited local transmission. This pattern is consistent with the notion that the enhanced immune evasion of Omicron lineages facilitated the local emergence and short-term spread of such recombinants. To further assess whether any of the identified NRLs subsequently contributed to larger transmission chains, all 21 NRL genomes were placed into the global SARS-CoV-2 phylogeny using UShER (Table S14). No evidence was found for expansion into major circulating recombinant lineages. Most genomes showed unique placements and clustered only with closely related contemporaneous background sequences. Declining incidence from July 2022 onwards (Figure S1), together with continued NPIs in October 2022, likely reduced transmission opportunities and constrained further spread. Following 2022, only a single additional recombinant lineage was detected (Fig. 2c), likely reflecting reduced incidence and thus fewer opportunities for co-infection and recombination [56], alongside declining sampling density in genomic surveillance systems [57]. In parallel, multiple recombinant lineages had already emerged and established global circulation during the late Omicron diversification phase [58]. Although the overall diversity of SARS-CoV-2 continued to increase, additional recombinant forms may have remained undetected owing to surveillance limitations, uneven sequencing coverage, and the often cryptic nature of recombination signals [24,47].

Across all NRLs, breakpoint hotspots were enriched in regions spanning ORF1b, Spike, and ORF3 to ORF7a, consistent with previous studies [29,47,59,60]. These regions encode proteins involved in viral replication and host–virus interactions, with variation in the spike protein in particular contributing to altered antigenicity and immune evasion underscoring their functional relevance in recombinant virus evolution. Interestingly, we also identified a putative breakpoint in ORF1a outside commonly reported regions in one NRL, observed independently in three cases (2022804576, 2022804577 and 2022804637) (Fig. 3). Although the corresponding breakpoint probability was elevated, it did not remain significant after FDR correction. Nevertheless, the recurrence of this breakpoint across independent recombinant genomes suggests that recombination is not restricted to established hotspot regions. This observation is consistent with recombination occurring across a broader genomic landscape, while sustained transmission may preferentially favor recombinant lineages carrying breakpoints within established hotspot regions such as Spike. Together, these results indicate that while recombination breakpoints are enriched in defined genomic hotspots, as also observed for CRLs (Fig. 3b), events in less frequently affected regions can likewise give rise to lineages capable of sustained transmission.

Methodological considerations and limitations

Despite the high-resolution nature of this regional genomic dataset, several methodological limitations should be considered. First, sequencing was performed using Oxford Nanopore Technologies with R9.4.1 flow cells. While this platform enables rapid and scalable viral genome generation and reliable consensus-level SARS-CoV-2 sequencing, it is associated with higher raw read error rates compared with

short-read technologies. Although consensus calling and read-level validation mitigate these limitations, residual uncertainty in low-frequency variation and breakpoint resolution cannot be fully excluded [61,62].

Second, the amplicon-based PCR strategy used for SARS-CoV-2 genomic surveillance may introduce biases. Uneven primer performance, amplicon dropout, and template switching during amplification can generate patterns resembling recombination signals. Although the ARTIC framework is extensively validated, such artefacts may affect coverage uniformity and complicate interpretation of complex genomic events, particularly in low viral load samples [61,63,64].

Third, co-infections were not systematically screened across the full dataset but inferred indirectly through recombinant genome detection using consensus-based screening. Therefore, their true prevalence is likely underestimated, limiting direct inference on the relationship between co-infection frequency and recombination dynamics. This limitation is particularly relevant during periods of high variant co-circulation, when transient co-infections may remain undetected without deep sequencing approaches [24,65–67]. Furthermore, our framework is optimized for detecting high-confidence, transmission-consistent recombinant genomes and is not designed to resolve transient intra-host recombination events described by deep sequencing studies. Resolving such within-host processes requires substantially higher sequencing depth and dedicated bioinformatic approaches beyond routine genomic surveillance and the present study [24].

Finally, despite stringent quality control and read-level validation, low-level contamination or index misassignment cannot be completely excluded in high-throughput amplicon sequencing workflows [68,69]. However, consistent breakpoint signals across multiple amplicon regions and supporting read-level evidence provide confidence in the identified recombinant genomes.

Conclusion

Our study provides a high-resolution view of SARS-CoV-2 evolution in a geographically confined region, revealing distinct lineage dynamics and a diverse yet highly transient landscape of recombinant genomes in northeastern Germany between 2022 and 2025. Despite substantial co-circulation of divergent lineages and direct evidence of co-infections, only a small subset of recombinant genomes showed signs of onward transmission, indicating that while recombination is frequently initiated, its epidemiological establishment remains uncommon. Recombination events were predominantly observed within Omicron-derived lineages and clustered in functionally relevant genomic regions, consistent with constrained but non-random evolutionary processes. The detection of multiple previously unrecognized recombinant genomes further highlights the extent viral diversity that remains unsolved by routine lineage assignment frameworks. At the same time, our results demonstrate that reliable detection of recombinant genomes requires analytical approaches that extend beyond consensus-based lineage assignment alone. While consensus genomes remain indispensable for large scale monitoring, they lack the resolution needed to robustly distinguish recombinant genomes from co-infections and technical artefacts. To address this, we implemented a recombination-focused two-stage

framework that combines high-sensitivity consensus-based screening (REBAR) with targeted raw-read validation. This approach enables efficient prioritization of candidate genomes and robust discrimination of true recombination events from co-infections and technical artefacts, without requiring exhaustive read-level analysis across entire datasets.

In conclusion, our findings demonstrate that regionally resolved genomic surveillance can uncover fine-scale evolutionary dynamics that remain obscured in global datasets, and that integrating sensitive screening with focused validation provides a practical and scalable strategy to capture recombinant SARS-CoV-2 diversity. Such recombination-aware analytical frameworks represent an important extension of current genomic monitoring systems, enabling more comprehensive assessment of viral evolution and supporting preparedness for emerging viral threats.

Methods

Sample collection, selection and quality criteria for SARS-CoV-2 analysis

Established in 2021, the CoMV-Gen Centre continued its core activities after its formal dissolution at the end of July 2023 through its integration into the Outbreak Competence Center (OCC-UMG), a joint structure of the Friedrich Loeffler Institute of Medical Microbiology at University Medicine Greifswald (FLI-UMG) and the Institute of Hygiene and Environmental Medicine (IHU). In this framework, a total of 4,675 SARS-CoV-2–positive samples from patients in M–V were collected by collaborating laboratories between 1 January 2022 and 11 January 2025 and subjected to whole-genome sequencing (WGS). The FLI-UMG served as the central coordinating site, being responsible for sequencing, bioinformatic processing, and phylogenetic analyses. For SARS-CoV-2 confirmation, RNA isolation, RT-PCR and variant-specific PCR, and/or sequencing analyses were performed in collaborating laboratories in MV (Dietrich-Bonhoeffer-Klinikum Neubrandenburg GmbH, Friedrich Loeffler-Institute of Medical Microbiology, University Medicine Greifswald, Friedrich-Loeffler-Institute Federal Research Institute for Animal Health, IMD Labor Greifswald GmbH, and Landesamt für Gesundheit und Soziales, Rostock).

Samples selected for WGS and downstream analyses met the following criteria: (i) availability of complete metadata, including sample collection date, SARS-CoV-2 diagnosis, and, where available, patient residence information based on ZIP code; and (ii) a RT-PCR cycle threshold (Ct) value of < 35. Only genomes with > 99% coverage were included in downstream bioinformatic and phylogenetic analyses. Samples, including most associated metadata and analytical results, are summarized in Table S11.

RNA extraction and real-time RT-PCR (qRT-PCR)

RNA was extracted from nasopharyngeal swab specimens using a range of commercial viral RNA extraction kits, including NucleoMag VET (Macherey-Nagel, Germany), QIAamp Viral RNA Mini Kit (QIAGEN, Germany), innuPREP DeltaPrep RNA Virus Plus Kit (Analytik Jena, Germany), STARMag 96 X4 Viral DNA/RNA 200 C (Seegene, South Korea), and/or TanBead automated extraction systems (TanBead, Taiwan), according to the manufacturers' instructions.

SARS-CoV-2 detection and/or variant screening was performed using multiple RT-PCR-based diagnostic assays, including the RealStar SARS-CoV-2 RT-PCR Kit (Altona Diagnostics, Germany), RIDAGENE SARS-CoV-2 Lineage Kit (r-biopharm, Germany), GSD NovaType II SARS-CoV-2 Kit (Gold Standard Diagnostics, Germany), Allplex SARS-CoV-2 Variant I and II assays, and Novaplex SARS-CoV-2 variant assays I, IV, and VII (Seegene, South Korea), as well as cobas SARS-CoV-2 assays (Roche, Switzerland), BD MAX SARS-CoV-2 and BD MAX SARS-CoV-2/Flu assays (Becton Dickinson, USA), CerTest VIASURE SARS-CoV-2 assays (CerTest Biotec, Spain), and/or Xpert Xpress SARS-CoV-2/Flu/RSV (Cepheid, USA).

SARS-CoV-2 whole-genome sequencing using Oxford Nanopore Technology

cDNA was synthesized from extracted total RNA using a multiplexed amplicon-based nanopore sequencing protocol as previously described (https://www.protocols.io/view/ncov-2019-sequencing-protocol-v3-locost-bh42j8ye). Amplicon pools were quantified using the Qubit dsDNA HS Assay Kit (Invitrogen, USA). For sequencing of Omicron variants and subvariants, an adapted primer scheme generating longer amplicons (V1200 Midnight) was used to ensure complete genome coverage. A total of 50 ng of amplicon pool per sample was end-repaired, barcoded, and ligated to sequencing adapters according to the manufacturer's protocol. Barcoded libraries were multiplexed (12 or 24 samples per run) and loaded onto R9.4.1 flow cells (Oxford Nanopore Technologies, UK) on MinION Mk1B or Mk1C devices.

Consensus sequence generation

Basecalling and initial processing were performed using the ARTIC bioinformatics pipeline (http://artic.network/ncov-2019). Reads were basecalled with Guppy v6.0.6 (high-accuracy model) and demultiplexed using the integrated Guppy barcoder. Reads were filtered by length (300–1,100 nt for ARTIC V3 and 250–2,500 nt for Midnight primer sets). Alignment to the SARS-CoV-2 reference genome (MN908947.3) was performed using minimap2 (https://github.com/lh3/minimap2). Primer sequences were removed, and variant calling and polishing were conducted with Medaka v1.6.0 (r941_min_high_g360) (https://github.com/nanoporetech/medaka). Consensus genomes were generated using bcftools (https://github.com/samtools/bcftools). Putative recombinant genomes were further validated and manually curated using Geneious Prime (v2025.2.2; GraphPad). Reads were re-

aligned with minimap2 to the reference genome, and consensus sequences were generated using a 60% majority base threshold based on quality-adjusted read support.

Phylogenetic analysis, bioinformatic processing, and data visualization

For the spatial analysis geographic coordinates were obtained via adaptive optical averaging using zip codes of patient data. Data were coarsened to ensure sufficient anonymization. Geographic geojson vector maps of the districts of M-V were created using open data (http://opendatalab.de/projects/geojson-utilities/) provided by the Federal Agency for Cartography and Geodesy (https://gdz.bkg.bund.de/). Visualization was performed using microreact (https://docs.microreact.org/)

The phylogenetic analysis and visualization were done as previously described [18]. A more stringent criteria was applied where only sequences with < 1% unsolved nucleotides where included. With the filtered sequences with additionally the Wuhan reference sequence were aligned leveraging MAFFT v7.525 [70] with disabled memory-saving mode to maximize accuracy. A maximum likelihood (ML) phylogenetic tree was inferred from the resulting multiple sequence alignment (MSA) using RaxML v8.2.12 (https://github.com/stamatak/standard-RAxML). The ML tree served as input for subsequent spatio-temporal analyses. Sampling dates were integrated for temporal phylogeny reconstruction using Augur refine tool. For visualization, the time tree was exported in JSON v2 format using Augur export and displayed with Auspice (Figure S4). Both Augur and Auspice are components of the Nextstrain toolkit [19].

For comparisons of circulating variants with global and European epidemiology, corresponding GISAID sample subsets were downloaded in April 2025 from Nextstrain (https://nextstrain.org/ncov) and filtered to match the analyzed time frame. Full lists are provided in Tables S12 (Europe) and S13 (global).

Recombination events were detected and visualized using rebar (v0.2.0) with a dataset of possible parent lineages from 2025-04-04. To satisfy the rebar prerequisite that the MSA should not contain insertions, a separate MSA was generated applying the augur align tool.

To extract the reads covering the variant-defining positions obtained from rebar, the mpileup command from bcftools (v.1.22-13) was used with simultaneous filtering for base quality (-Q 15) and skipping indels. For each position, the number of reads supporting the corresponding putative donors/parents was recorded and subsequently used to generate raw read plots in both absolute and relative terms. Private mutations, defined as mutations that could not be assigned to any of the parental variants, were excluded. Expert input was then used to set/establish a system of thresholds for the final sample classification. First, samples were classified as coinfections when more than 25% of all defining positions did not show a “dominant” parental variant, meaning that at these positions neither parental base was supported by at least 90% of the reads. Furthermore, recombinant classification required that

each parental variant was supported by at least three consecutive bases with ≥ 90% read support and a minimum read depth 40. Samples not meeting these criteria were considered co-infections.

Statistical analyses of breakpoint distributions

The genomic breakpoint position of recombinants cannot be determined exactly because the genomic sequences between the last defining position of one parental lineage and the first defining position of the other parental lineage are identical in both parents. Consequently, rebar reports, for each sample, the genomic interval containing the putative recombination breakpoint. To statistically assess whether these breakpoint intervals are distributed differently across SARS-CoV-2 genes between NRLs and CRLs, overlaps between breakpoint intervals and gene coordinates were counted. Each gene was then analyzed using Fisher's exact test, followed by multiple-testing correction with the Benjamini–Hochberg procedure.

Details of the permutation test are provided in the Supplementary Methods.

Data availability

The analyzed sequences created are accessible under the EBI-ENA umbrella project PRJEB115949. GISAID sequences used in this study are accessible are listen in tables S12 and S13.

Table 1
Comparison of gene–breakpoint interval overlaps in novel recombinant lineages (NRLs) and circulating recombinant lineages (CRLs).

Gene	Count NRLs	Count CRLs	Fraction NRLs(%)	Fraction CRLs(%)	p-value[1]	FDR[1]
ORF1a	8	41	11.11	7.24	0.241316	0.389414
ORF1b	14	58	19.44	10.25	0.028340	0.113361
S	23	334	31.94	59.01	0.000015	0.000120
ORF3a	6	31	8.33	5.48	0.292060	0.389414
E	6	30	8.33	5.3	0.279553	0.389414
M	6	24	8.33	4.24	0.135194	0.360516
ORF6	5	24	6.94	4.24	0.359690	0.411074
ORF7a	4	24	5.56	4.24	0.544140	0.544140

[1] Statistical significance was assessed using Fisher's exact test, followed by multiple-testing correction using the Benjamini–Hochberg procedure. P-values and false discovery rates (FDRs) range from 0 to 1.

Declarations

Ethical approval

The project “CoMV-Gen: SARS-CoV-2 in Mecklenburg-Vorpommern: Genetische Analyse und Nachverfolgung” was approved by the ethics committee of the University of Greifswald, Germany (BB 125/21). All data have been aggregated and anonymized for analyses.

Competing interests

The authors declare no competing interests.

Funding

This study was conducted within the framework of the CoMV-Gen project and was funded by the Ministry of Social Affairs, Health and Sports of Mecklenburg-Vorpommern (MV-Schutzfonds, grant number: 0680 894.01 MG 80) until July 31, 2023. Subsequent work was carried out without external funding. The funders had no role in the study design, data collection, analysis, interpretation, or manuscript preparation.

Author contributions

A.T., A.P., K.B. and C.K. conceived the study. A.T., A.P., and C.K. performed the analysis, validation and interpretation of the data and contributed to the writing of the original draft. K.B. and N.O.H. were responsible for project administration and resources. K.L., T.S. and V.B. contributed to the acquisition of the samples. All authors contributed to the writing of the manuscript and read and approved the final version of the manuscript.

Acknowledgements

We gratefully acknowledge all sample and data contributors. We further thank the technicians of the FLI-UMG, in particular A. Kohler and C. Wiede, for their excellent support and valuable contributions to this study. ChatGPT (OpenAI) was used to improve the language and readability of the manuscript. All scientific content, interpretations, and conclusions were developed and verified by the authors.

References

1. Zhu N et al (2020) A Novel Coronavirus from Patients with Pneumonia in China, 2019. N Engl J Med 382:727–733. https://doi.org/10.1056/NEJMoa2001017
2. Saravanan KA et al (2022) Role of genomics in combating COVID-19 pandemic. Gene 823:146387. https://doi.org/10.1016/j.gene.2022.146387
3. Shiraz R, Tripathi S (2023) Enhanced recombination among Omicron subvariants of SARS-CoV-2 contributes to viral immune escape. J Med Virol 95:e28519. https://doi.org/10.1002/jmv.28519
4. Tegally H et al (2021) Detection of a SARS-CoV-2 variant of concern in South Africa. Nature 592:438–443. https://doi.org/10.1038/s41586-021-03402-9
5. Faria NR et al (2021) Genomics and epidemiology of the P.1 SARS-CoV-2 lineage in Manaus, Brazil. Science 372:815–821. https://doi.org/10.1126/science.abh2644
6. Volz E et al (2021) Assessing transmissibility of SARS-CoV-2 lineage B.1.1.7 in England. Nature 593:266–269. https://doi.org/10.1038/s41586-021-03470-x
7. Twohig KA et al (2022) Hospital admission and emergency care attendance risk for SARS-CoV-2 delta (B.1.617.2) compared with alpha (B.1.1.7) variants of concern: a cohort study. Lancet Infect Dis 22:35–42. https://doi.org/10.1016/S1473-3099(21)00475-8
8. Mohsin M, Mahmud S, Omicron (2022) SARS-CoV-2 variant of concern: A review on its transmissibility, immune evasion, reinfection, and severity. Med (Baltim) 101:e29165. https://doi.org/10.1097/MD.0000000000029165
9. Ward IL et al (2022) Risk of covid-19 related deaths for SARS-CoV-2 omicron (B.1.1.529) compared with delta (B.1.617.2): retrospective cohort study. BMJ 378:e070695. https://doi.org/10.1136/bmj-2022-070695
10. Perez-Guzman PN et al (2023) Epidemiological drivers of transmissibility and severity of SARS-CoV-2 in England. Nat Commun 14:4279. https://doi.org/10.1038/s41467-023-39661-5
11. Lewnard JA et al (2022) Clinical outcomes associated with SARS-CoV-2 Omicron (B.1.1.529) variant and BA.1/BA.1.1 or BA.2 subvariant infection in Southern California. Nat Med 28:1933–1943. https://doi.org/10.1038/s41591-022-01887-z
12. Lauring AS et al (2022) Clinical severity of, and effectiveness of mRNA vaccines against, covid-19 from omicron, delta, and alpha SARS-CoV-2 variants in the United States: prospective observational study. BMJ 376:e069761. https://doi.org/10.1136/bmj-2021-069761
13. Wang X, Zein J, Ji X, Lin DY (2023) Impact of Vaccination, Prior Infection, and Therapy on Omicron Infection and Mortality. J Infect Dis 227:970–976. https://doi.org/10.1093/infdis/jiac460
14. Lai MM (1992) Genetic recombination in RNA viruses. Curr Top Microbiol Immunol 176:21–32. https://doi.org/10.1007/978-3-642-77011-1_2
15. Bonnet J, Fraile A, Sacristan S, Malpica JM, Garcia-Arenal F (2005) Role of recombination in the evolution of natural populations of Cucumber mosaic virus, a tripartite RNA plant virus. Virology 332:359–368. https://doi.org/10.1016/j.virol.2004.11.017

16. Varsani A, Lefeuvre P, Roumagnac P, Martin D (2018) Notes on recombination and reassortment in multipartite/segmented viruses. Curr Opin Virol 33:156–166. https://doi.org/10.1016/j.coviro.2018.08.013
17. Eales O et al (2024) Key Challenges for Respiratory Virus Surveillance while Transitioning out of Acute Phase of COVID-19 Pandemic. Emerg Infect Dis 30. https://doi.org/10.3201/eid3002.230768
18. Kohler C et al (2023) Neighbourhood watch: genomic epidemiology of SARS-CoV-2 variants circulating in a German federal state, Mecklenburg-Western Pomerania, in 2020–2022. Emerg Microbes Infect 12:2245916. https://doi.org/10.1080/22221751.2023.2245916
19. Hadfield J et al (2018) Nextstrain: real-time tracking of pathogen evolution. Bioinformatics 34:4121–4123. https://doi.org/10.1093/bioinformatics/bty407
20. Rambaut A et al (2020) A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. Nat Microbiol 5:1403–1407. https://doi.org/10.1038/s41564-020-0770-5
21. O'Toole A et al (2021) Assignment of epidemiological lineages in an emerging pandemic using the pangolin tool. Virus Evol 7:veab064. https://doi.org/10.1093/ve/veab064
22. Shu Y, McCauley JGISAID (2017) Global initiative on sharing all influenza data - from vision to reality. Euro Surveill 22. https://doi.org/10.2807/1560-7917.ES.2017.22.13.30494
23. Benson DA et al (2013) GenBank. Nucleic Acids Res 41:D36–42. https://doi.org/10.1093/nar/gks1195
24. Pipek OA et al (2024) Systematic detection of co-infection and intra-host recombination in more than 2 million global SARS-CoV-2 samples. Nat Commun 15:517. https://doi.org/10.1038/s41467-023-43391-z
25. Bolze A et al (2022) Evidence for SARS-CoV-2 Delta and Omicron co-infections and recombination. *Med* 3, 848–859 e844 https://doi.org/10.1016/j.medj.2022.10.002
26. Leinonen R et al (2011) The European Nucleotide Archive. Nucleic Acids Res 39:D28–31. https://doi.org/10.1093/nar/gkq967
27. Brito AF et al (2022) Global disparities in SARS-CoV-2 genomic surveillance. Nat Commun 13:7003. https://doi.org/10.1038/s41467-022-33713-y
28. VanInsberghe D, Neish AS, Lowen AC, Koelle K (2021) Recombinant SARS-CoV-2 genomes circulated at low levels over the first year of the pandemic. Virus Evol 7:veab059. https://doi.org/10.1093/ve/veab059
29. Li JY et al (2024) Comprehensive detection and dissection of interlineage recombination events in the SARS-CoV-2 pandemic. Virus Evol 10:veae074. https://doi.org/10.1093/ve/veae074
30. Zhan SH et al (2025) A Pandemic-Scale Ancestral Recombination Graph for SARS-CoV-2. *bioRxiv* https://doi.org/10.1101/2023.06.08.544212
31. Choga WT et al (2024) Genomic epidemiology and immune escape of SARS-CoV-2 recombinant strains circulating in Botswana. IJID Reg 13:100484. https://doi.org/10.1016/j.ijregi.2024.100484

32. Brauner JM et al (2021) Inferring the effectiveness of government interventions against COVID-19. Science 371. https://doi.org/10.1126/science.abd9338
33. Nouvellet P et al (2021) Reduction in mobility and COVID-19 transmission. Nat Commun 12:1090. https://doi.org/10.1038/s41467-021-21358-2
34. Flaxman S et al (2020) Estimating the effects of non-pharmaceutical interventions on COVID-19 in Europe. Nature 584:257–261. https://doi.org/10.1038/s41586-020-2405-7
35. Zebrowski P, Boyarshinov G, Odintsova A, Rovenskaya E, Explaining (2025) COVID-19 dynamics through user activity data from digital platforms with Yandex's self-isolation index as a case study. Sci Rep 15:40709. https://doi.org/10.1038/s41598-025-24240-z
36. Iezadi S et al (2021) Effectiveness of non-pharmaceutical public health interventions against COVID-19: A systematic review and meta-analysis. PLoS ONE 16:e0260371. https://doi.org/10.1371/journal.pone.0260371
37. Harder T et al (2021) Effectiveness of COVID-19 vaccines against SARS-CoV-2 infection with the Delta (B.1.617.2) variant: second interim results of a living systematic review and meta-analysis, 1 January to 25 August 2021. *Euro Surveill* 26 https://doi.org/10.2807/1560-7917.ES.2021.26.41.2100920
38. Andrews N et al (2022) Covid-19 Vaccine Effectiveness against the Omicron (B.1.1.529) Variant. N Engl J Med 386:1532–1546. https://doi.org/10.1056/NEJMoa2119451
39. Goldberg Y et al (2021) Waning Immunity after the BNT162b2 Vaccine in Israel. N Engl J Med 385:e85. https://doi.org/10.1056/NEJMoa2114228
40. Han AX et al (2023) SARS-CoV-2 diagnostic testing rates determine the sensitivity of genomic surveillance programs. Nat Genet 55:26–33. https://doi.org/10.1038/s41588-022-01267-w
41. Oh DY et al (2025) A public health perspective of SARS-CoV-2 evolution and surveillance strategies in Germany from 2020 to 2023. Commun Med (Lond) 5:468. https://doi.org/10.1038/s43856-025-01093-1
42. Zeghbib S, Kemenesi G, Jakab F (2023) The importance of equally accessible genomic surveillance in the age of pandemics. Biol Futur 74:81–89. https://doi.org/10.1007/s42977-023-00164-5
43. Oltean HN et al (2023) Sentinel Surveillance System Implementation and Evaluation for SARS-CoV-2 Genomic Data, Washington, USA, 2020–2021. Emerg Infect Dis 29:242–251. https://doi.org/10.3201/eid2902.221482
44. Lam C et al (2023) A laboratory framework for ongoing optimization of amplification-based genomic surveillance programs. Microbiol Spectr 11:e0220223. https://doi.org/10.1128/spectrum.02202-23
45. Silva T et al (2024) Enhancing the epidemiological surveillance of SARS-CoV-2 using Sanger sequencing to identify circulating variants and recombinants. Braz J Microbiol 55:2085–2099. https://doi.org/10.1007/s42770-024-01387-x
46. Jung A et al (2024) Tracking the prevalence and emergence of SARS-CoV-2 variants of concern using a regional genomic surveillance program. Microbiol Spectr 12:e0422523. https://doi.org/10.1128/spectrum.04225-23

47. Turakhia Y et al (2022) Pandemic-scale phylogenomics reveals the SARS-CoV-2 recombination landscape. Nature 609:994–997. https://doi.org/10.1038/s41586-022-05189-9
48. Wertheim JO et al (2022) Detection of SARS-CoV-2 intra-host recombination during superinfection with Alpha and Epsilon variants in New York City. Nat Commun 13:3645. https://doi.org/10.1038/s41467-022-31247-x
49. Piccoli L et al (2020) Mapping Neutralizing and Immunodominant Sites on the SARS-CoV-2 Spike Receptor-Binding Domain by Structure-Guided High-Resolution Serology. *Cell* 183, 1024–1042 e1021 https://doi.org/10.1016/j.cell.2020.09.037
50. Liu L et al (2020) Potent neutralizing antibodies against multiple epitopes on SARS-CoV-2 spike. Nature 584:450–456. https://doi.org/10.1038/s41586-020-2571-7
51. Huo J et al (2020) Neutralizing nanobodies bind SARS-CoV-2 spike RBD and block interaction with ACE2. Nat Struct Mol Biol 27:846–854. https://doi.org/10.1038/s41594-020-0469-6
52. Gao Y et al (2020) Structure of the RNA-dependent RNA polymerase from COVID-19 virus. Science 368:779–782. https://doi.org/10.1126/science.abb7498
53. Hillen HS et al (2020) Structure of replicating SARS-CoV-2 polymerase. Nature 584:154–156. https://doi.org/10.1038/s41586-020-2368-8
54. Gribble J et al (2021) The coronavirus proofreading exoribonuclease mediates extensive viral recombination. PLoS Pathog 17:e1009226. https://doi.org/10.1371/journal.ppat.1009226
55. Tamura T et al (2023) Virological characteristics of the SARS-CoV-2 XBB variant derived from recombination of two Omicron subvariants. Nat Commun 14:2800. https://doi.org/10.1038/s41467-023-38435-3
56. Ma KC et al (2024) Genomic Surveillance for SARS-CoV-2 Variants: Circulation of Omicron XBB and JN.1 Lineages - United States, May 2023-September 2024. MMWR Morb Mortal Wkly Rep 73:938–945. https://doi.org/10.15585/mmwr.mm7342a1
57. Gu H et al (2026) Optimizing global genomic surveillance for early detection of emerging SARS-CoV-2 variants. Nat Commun. https://doi.org/10.1038/s41467-026-70664-0
58. Roemer C et al (2023) SARS-CoV-2 evolution in the Omicron era. Nat Microbiol 8:1952–1959. https://doi.org/10.1038/s41564-023-01504-w
59. Harvey WT et al (2021) SARS-CoV-2 variants, spike mutations and immune escape. Nat Rev Microbiol 19:409–424. https://doi.org/10.1038/s41579-021-00573-0
60. Focosi D, Maggi F (2022) Recombination in Coronaviruses, with a Focus on SARS-CoV-2. *Viruses* 14 https://doi.org/10.3390/v14061239
61. Quick J et al (2017) Multiplex PCR method for MinION and Illumina sequencing of Zika and other virus genomes directly from clinical samples. Nat Protoc 12:1261–1276. https://doi.org/10.1038/nprot.2017.066
62. Bull RA et al (2020) Analytical validity of nanopore sequencing for rapid SARS-CoV-2 genome analysis. Nat Commun 11:6272. https://doi.org/10.1038/s41467-020-20075-6

63. Lambisia AW et al (2022) Optimization of the SARS-CoV-2 ARTIC Network V4 Primers and Whole Genome Sequencing Protocol. Front Med (Lausanne) 9:836728. https://doi.org/10.3389/fmed.2022.836728
64. Brakenhoff RH, Schoenmakers JG, Lubsen NH (1949) (1991) Chimeric cDNA clones: a novel PCR artifact. *Nucleic Acids Res* 19. https://doi.org/10.1093/nar/19.8.1949
65. Rockett RJ et al (2022) Co-infection with SARS-CoV-2 Omicron and Delta variants revealed by genomic surveillance. Nat Commun 13:2745. https://doi.org/10.1038/s41467-022-30518-x
66. Bal A et al (2022) Detection and prevalence of SARS-CoV-2 co-infections during the Omicron variant circulation in France. Nat Commun 13:6316. https://doi.org/10.1038/s41467-022-33910-9
67. Lythgoe KA et al (2021) SARS-CoV-2 within-host diversity and transmission. Science 372. https://doi.org/10.1126/science.abg0821
68. Mwangi P et al (2022) A decontamination strategy for resolving SARS-CoV-2 amplicon contamination in a next-generation sequencing laboratory. Arch Virol 167:1175–1179. https://doi.org/10.1007/s00705-022-05411-z
69. Lagerborg KA et al (2022) Synthetic DNA spike-ins (SDSIs) enable sample tracking and detection of inter-sample contamination in SARS-CoV-2 sequencing workflows. Nat Microbiol 7:108–119. https://doi.org/10.1038/s41564-021-01019-2
70. Katoh K, Standley DM (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol 30:772–780. https://doi.org/10.1093/molbev/mst010

Figures

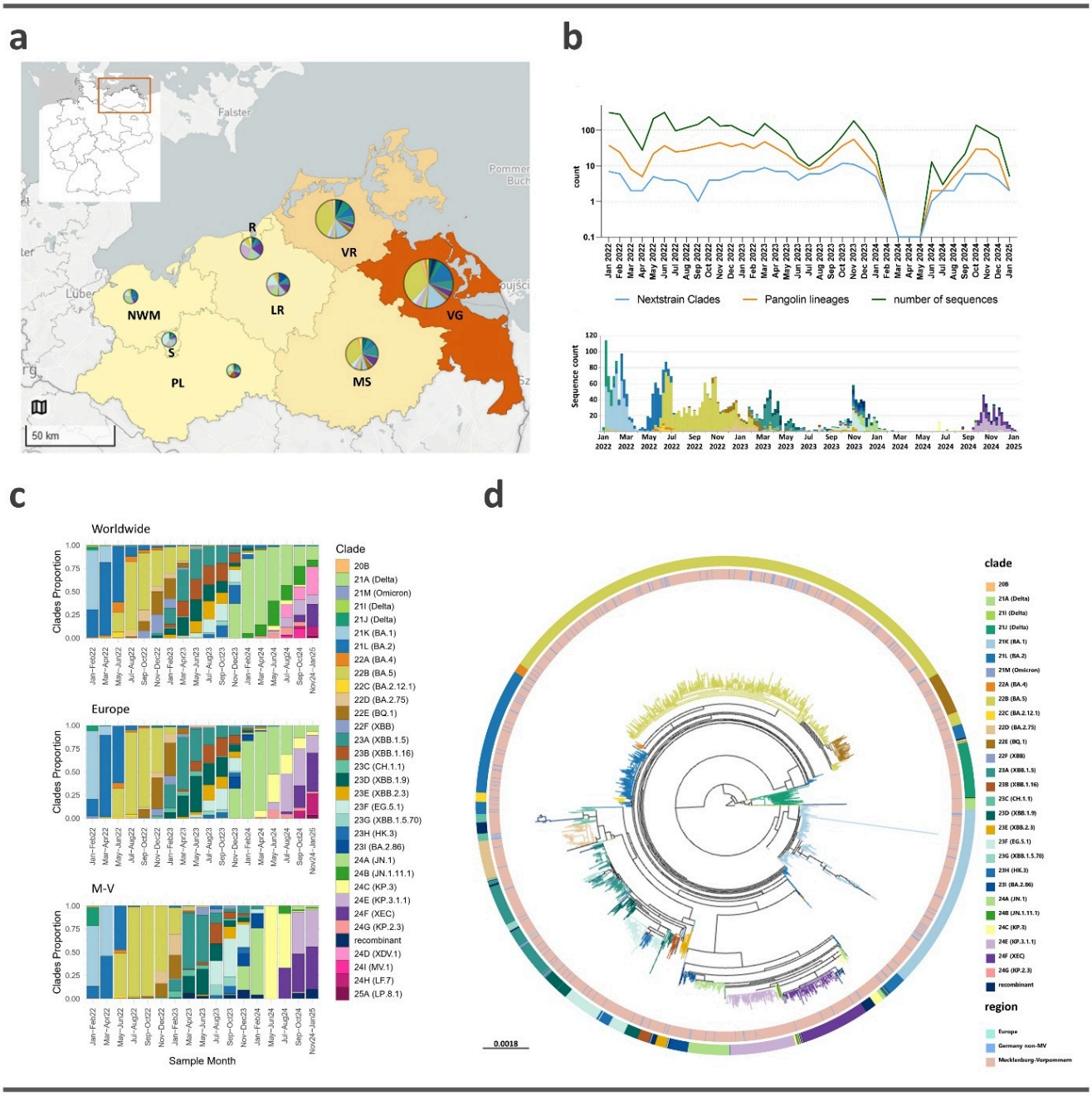


Figure 1

SARS-CoV-2 clade dynamics, diversity and phylogeny in Mecklenburg–Western Pomerania (M–V), Germany. **a** Distribution of SARS-CoV-2 clades across districts in M–V (January 2022–January 2025; n = 2,930), including Vorpommern-Rügen (VG, n = 464), Vorpommern-Greifswald (VR, n = 2,142), Rostock (R, n = 45), Landkreis Rostock (LR, n = 49), Mecklenburgische Seenplatte (MS, n = 188), Nordwestmecklenburg (NWM, n = 14), Schwerin (S, n = 15) and Ludwigslust-Parchim (PL, n = 13). Sequences lacking regional metadata or assigned outside M–V (n = 430; 219 external, 211 missing) were excluded. Recombinant Pango lineages (n = 76) were grouped into a single “recombinant” clade.

The clade colour legend for a, b and d is shown in panel d. **b** Temporal dynamics of SARS-CoV-2 diversity in M–V (January 2022–January 2025; n = 3,360). Monthly sampling intensity (dark green) and diversity of Nextstrain clades (blue) and Pangolin lineages (orange) are shown (top). Weekly clade distributions illustrate lineage turnover and shifting dominance over time (bottom). No samples were available between March and May 2024. **c**Comparison of circulating SARS-CoV-2 clades in M–V (n = 3,360) with European (n = 2,150) and global (n = 1,883) datasets (January 2022–January 2025), based on randomly subsampled sequences from GISAID (see Tables S11, S12 and S13). **d** Maximum-likelihood phylogeny of 3,360 SARS-CoV-2 genomes from M–V with contextual sequences from Germany and Europe. The tree was inferred using RAxML v8.2.12 with the Wuhan reference genome (MN908947.3; clade 19A) as outgroup. Outer and inner rings indicate Nextstrain clade assignment and geographic origin, respectively. Clade assignments were performed using Nextclade and lineage assignments using Pangolin. Recombinant lineages (21 Pango lineages; n = 76 sequences), including XAK, XAL, XBC.1.3.1, XBF, XBL, XBS, XBW, XCH.1, XCT, XCT.1, XDA, XDD, XDQ, XDY, XE, XEK, XEK.1, XEK.2, XEL, XEU and XM, were collapsed into a single category recombinant.

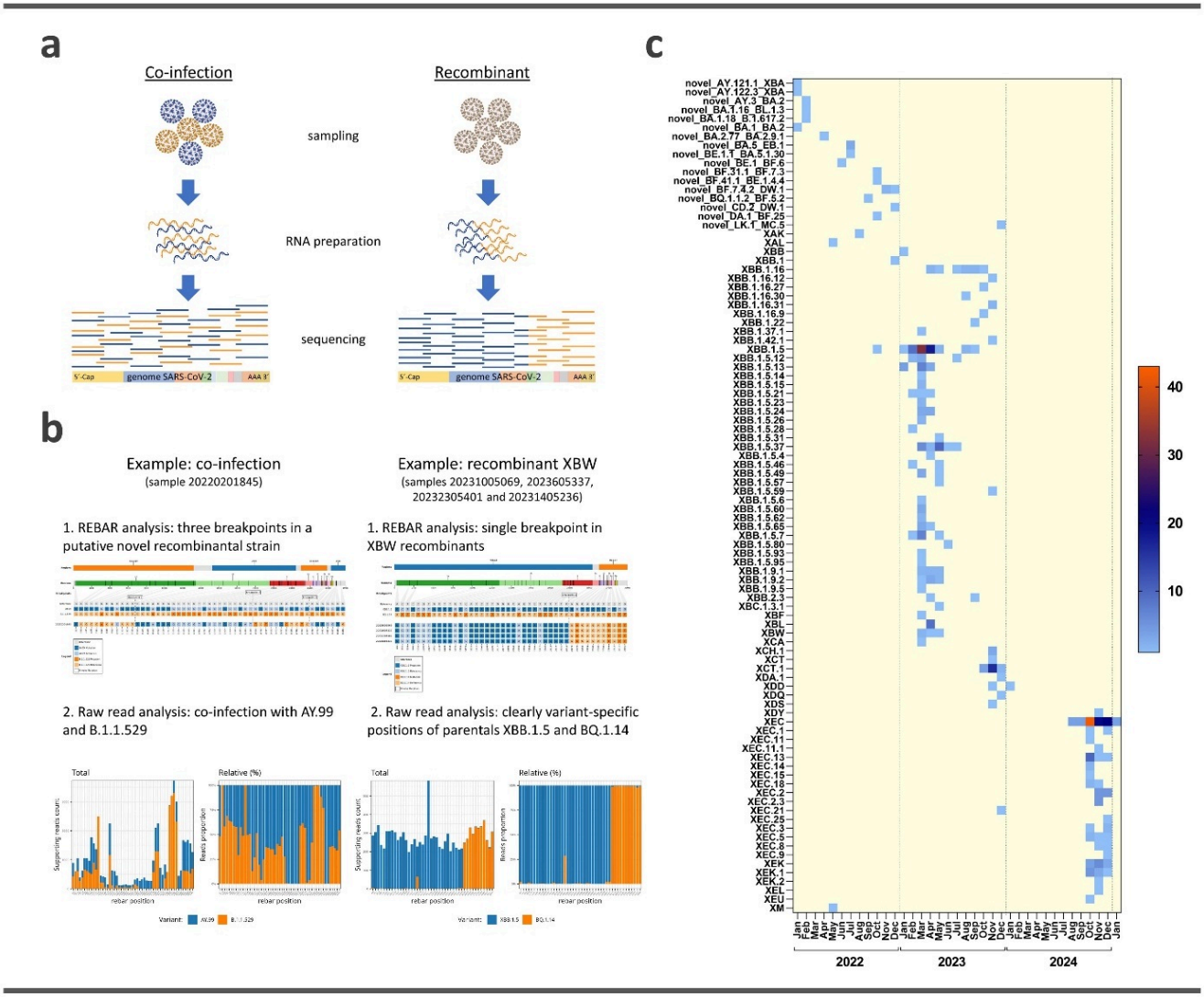


Figure 2

Identification and characterization of SARS-CoV-2 recombinants. **a** Schematic representation of co-infections and true recombinants at the virus, RNA and sequencing levels, illustrating the conceptual distinction between co-infection and recombinant genomes. **b**Representative examples of co-infections and recombinants. Putative recombinant candidates were initially identified using REBAR and subsequently validated or rejected by raw-read–level analysis. True recombinants exhibit well-defined breakpoints and are characterized by clearly delineated genomic segments with predominant (>90%) contributions from the respective parental lineages (right), without substantial admixture from other lineages across extended genomic regions (left). **c** Temporal distribution of REBAR- and Nextstrain/Pangolin-assigned recombinant SARS-CoV-2 lineages in M–V (January 2022–January 2025). In total, 431 sequences were assigned to 94 distinct putative recombinant lineages. Color intensity indicates the number of sequences per lineage (y-axis) over time (x-axis), with light yellow indicating absence of recombinant sequences.

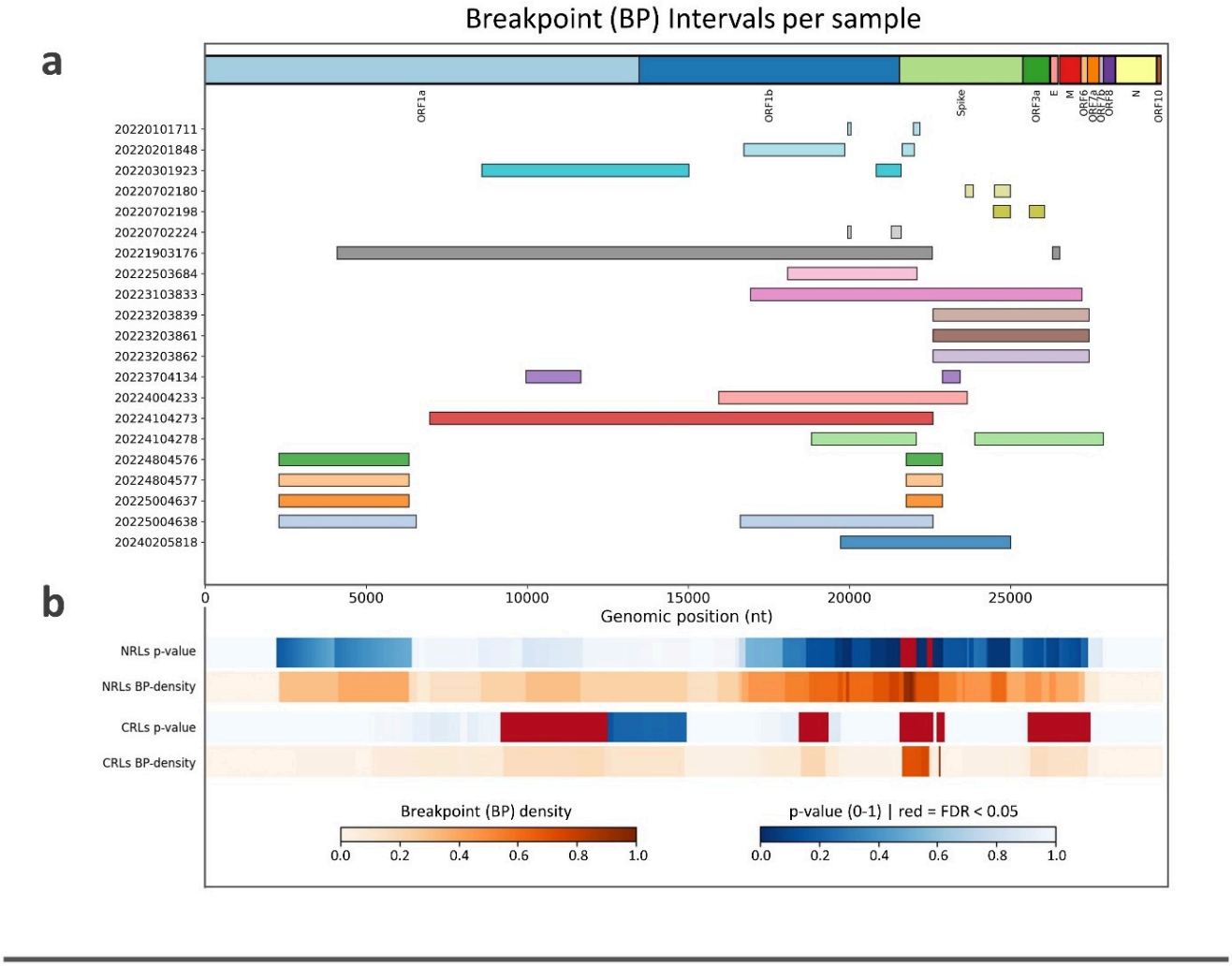


Figure 3

Breakpoint intervals and enrichment analysis in recombinant SARS-CoV-2 lineages. **a**Stacked breakpoint (BP) intervals across 21 newly identified recombinant sequences (NRLs), aligned to the SARS-CoV-2

reference genome (top), enabling direct comparison of breakpoint positions. The reference genome and genomic coordinate axis are shared between panels a and b. **b** Breakpoint (BP) distributions in circulating and novel recombinant lineages (CRLs and NRLs) were analyzed using a length-preserving permutation framework. Empirical p-values were computed genome-wide, aggregated at base resolution, and adjusted for multiple testing using the Benjamini–Hochberg false discovery rate (FDR). BP density is shown in orange, and permutation-derived p-values in blue. Regions with FDR < 0.05 are highlighted in dark red.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementalTables114.xlsx
- SupplementalFiguresMethods.pdf