Virologica Sinica 41 (2026) 382–391

Contents lists available at ScienceDirect

Virologica Sinica

journal homepage: www.keaipublishing.com/en/journals/virologica-sinica
www.virosin.org

Research Article

Genomic-epidemiological analysis of 15 million SARS-CoV-2 genomes reveals accelerated fitness gain of JN.1 lineage

Zhong-yi Lei [a,b], Qian-tong Jin [c,d], Xiao-min Zhang [b], Xiao-chen Bo [b], Zi-lin Ren [c,d,*], Yi-gang Tong [a,*], Ming Ni [b,*]

[a] College of Life Science and Technology, Beijing University of Chemical Technology, Beijing, 100029, China
[b] Advanced & Interdisciplinary Biotechnology, Academy of Military Medical Sciences, Beijing, 100850, China
[c] Changchun Veterinary Research Institute, Chinese Academy of Agricultural Sciences, State Key Laboratory of Pathogen and Biosecurity, Key Laboratory of Jilin Province for Zoonosis Prevention and Control, Changchun, 130122, China
[d] School of Information Science and Technology, Northeast Normal University, Changchun, 130117, China

ARTICLE INFO

Keywords:
SARS-CoV-2 evolution
JN.1
Genomic epidemiology
Lineage fitness
Physicochemical and structural feature

ABSTRACT

The evolution of SARS-CoV-2 has been driven by successive globally circulating waves, including the Alpha and Delta lineages, early Omicron (BA.1–BA.5), XBB, and the recently dominant JN.1 lineages. Although the marked advantage in fitness of early Omicron over Delta lineages has been recognized, there is a lack of systematic evaluation of SARS-CoV-2 fitness across 2020 to 2025. Here, we analyzed 15.23 million SARS-CoV-2 genomes available through May 2025. The accumulation of mutations in the spike protein of the virus has continued to accelerate over time, whereas the trend slowed in the other viral proteins. Using a Bayesian genomic-epidemiological framework, we estimated that lineage fitness increased approximately linearly from 2021 to 2025. Notably, JN.1 lineages exhibited a significantly higher rate of fitness gain than their predecessor XBB and earlier Omicron lineages. We further analyzed characteristic mutations of JN.1 and found that those in the receptor-binding domain were associated with larger alterations in residue hydropathy, charge, and structural surface exposure relative to other lineages. These findings suggest JN.1 as a distinct evolutionary stage and underscore the importance of sustained genomic surveillance.

INTRODUCTION

Since December 2019, SARS-CoV-2 has continued to circulate globally, causing recurrent waves of infection. Over five years of sustained human-to-human transmission, the virus has undergone extensive evolution, generating genetically diverse lineages (Li et al., 2021, 2024; Markov et al., 2023). The World Health Organization (WHO) classifies biologically significant lineages as Variants of Concern (VOCs), Variants of Interest (VOIs), and Variants Under Monitoring (VUMs). Major VOCs include the Alpha and Delta, and in November 2021, the Omicron sublineages BA.1 and BA.2 emerged, exhibiting substantial genomic variation concentrated in the spike (S) protein (Harvey et al., 2021; Ma et al., 2023). Subsequently, the recombinant XBB lineages dominated global circulation from late 2022 to October 2023, and were then largely replaced by JN.1 from November 2023 onward (Jian et al., 2025; Yajima et al., 2024; Yang et al., 2024). The JN.1 gave rise to multiple globally circulating sublineages, including KP.3 in 2024 and LP.8 from July 2024 to May 2025, and these have been associated with further antigenic escape and reduced protection from vaccination (Ioannou et al., 2025; Wang et al., 2024; Xu et al., 2024). Notably, although the circulation and evolution of SARS-CoV-2 continuously persist in human populations, genomic surveillance of SARS-CoV-2 has declined substantially since the WHO declared the end of the public health emergency in May 2023. In the subsequent three years (2023–2025), weekly submissions to GISAID decreased from ~131,000 to ~3,160, limiting timely and comprehensive assessment of ongoing evolutionary change.

Quantifying the fitness of SARS-CoV-2 lineages over the past five years is important yet remains relatively understudied. Recent studies

* Corresponding authors.
E-mail addresses: niming@bmi.ac.cn (M. Ni), tongyigang@mail.buct.edu.cn (Y. Tong), zilin.ren@outlook.com (Z. Ren).
Peer review under the responsibility of editorial board of *Virologica Sinica*.

https://doi.org/10.1016/j.virs.2026.03.015
Received 15 January 2026; Accepted 30 March 2026
Available online 8 April 2026

have primarily focused on the antigenic and virological characteristics of individual novel variants, such as LP.8.1 and NB.1.8 (Guo et al., 2025; Liu et al., 2025). However, a comprehensive, population-level analysis that jointly characterizes long-term genomic variation and lineage fitness across the full course of SARS-CoV-2 circulation is still lacking, and is essential for an integrated view of how viral evolution has unfolded and interacted with the context of changing host immunity.

Prior studies have well-documented the rapid global replacement of preceding variants by Omicron lineages in late 2021 (Chaguza et al., 2022; Paton et al., 2022; Viana et al., 2022). Obermeyer et al. (2022) integrated global genomic and epidemiological data, demonstrating markedly increased fitness of Omicron BA.1 and BA.2 relative to earlier VOCs. However, it remains unclear whether the rate of fitness gain has further accelerated in the past three years, particularly following the emergence of XBB and JN.1, which subsequently triggered large waves of global transmission.

In this study, we analyzed amino-acid-altering mutations across 15.23 million SARS-CoV-2 genomes collected between December 2019 and May 2025. Using a genomic epidemiological approach, we found that the emergence of JN.1, but not XBB, was associated with a significant acceleration in fitness gain relative to preceding XBB and earlier Omicron lineages. Compared with XBB and early Omicron variants, JN.1 exhibited more frequent recurrent and shifted mutations at previously mutated sites, along with greater variation in residual physicochemical properties and surface exposure. These results underscore the importance of sustained genomic surveillance and continued functional investigation.

RESULTS

Mutation accumulation of SARS-CoV-2 from 2019 to 2025

From the prototypic Wuhan-Hu-1 strain (GISAID accession EPI_ISL_402124) in late 2019, more than 16 million SARS-CoV-2 genomes have been submitted to GISAID (averaging 258,370 submissions per month from January 2020 to May 2025, Fig. 1A). This genomic surveillance dataset for a single species enables systematic, population-level genomic and evolutionary analyses. However, it should be noted that the submissions declined sharply after May 2025, falling to <1000 per week.

SARS-CoV-2 variants were assigned to Pango lineages (Rambaut et al., 2020) based on their genomic mutations. Using one-week time bins, we quantified the number of amino-acid-altering mutations in lineages during the five years, (see thresholds in Methods, Fig. 1B). The mutations of SARS-CoV-2 lineages accumulate over time and, notably, progress through discrete linear phases. Specifically, the overall linear increase was punctuated by three major leaps over the past five years, coinciding with the emergence of Alpha, Omicron, and JN.1 variants (Fig. 1B). The most substantial increase was accompanied by the sudden emergence of Omicron in late 2021. Subsequent XBB sublineages contributed only a modest, rather than dramatic increase in mutation counts.

Acceleration of mutation accumulation in the S protein

The accumulation of mutations was largely attributable to those in the S protein, rather than in other proteins (Fig. 1C and D; Supplementary Fig. S1–4). Changes in the mutational spectrum of S protein are shown in Fig. 1E, which reveals dramatic mutation shifts induced by the emergence of Alpha, Omicron, XBB, and JN.1 variants.

SARS-CoV-2 reached its highest accumulation rate of mutations at the genome-wide level during the pre-Omicron VOCs periods, with a rate of 0.25 amino acid (AA) mutations per week, followed by a subsequent decline (Fig. 1B). By contrast, the S protein exhibited a divergent trend: its accumulation rate increased markedly and continuously from the emergence of Omicron onward. In particular, the JN.1 variants accumulated significantly faster than XBB, reaching 0.128 AA mutations per week (Fig. 1F). An opposite trend was observed for the other proteins, where the weekly accumulation rate in JN.1 variants was as low as 0.052 (Fig. 1G). These results suggest the ongoing adaptation of SARS-CoV-2 is becoming increasingly predominant in the S protein.

JN.1 exhibited acceleration in fitness increasing

The mutations exhibited three major leaps (consistent with the emergence of Alpha, Omicron, and JN.1), and we next compared the variation of viral fitness for lineages across these leap-associated periods. It is challenging and infeasible to integrate the effects of separate mutations to measure the fitness, considering the highly diverse functionalities and epistasis. Therefore, we adopted a genomic epidemiology approach to estimate viral fitness at the lineage level. Namely, the fitness of lineages was inferred by analyzing their prevalence dynamics in humans, which accounted for the competition among co-circulating lineages.

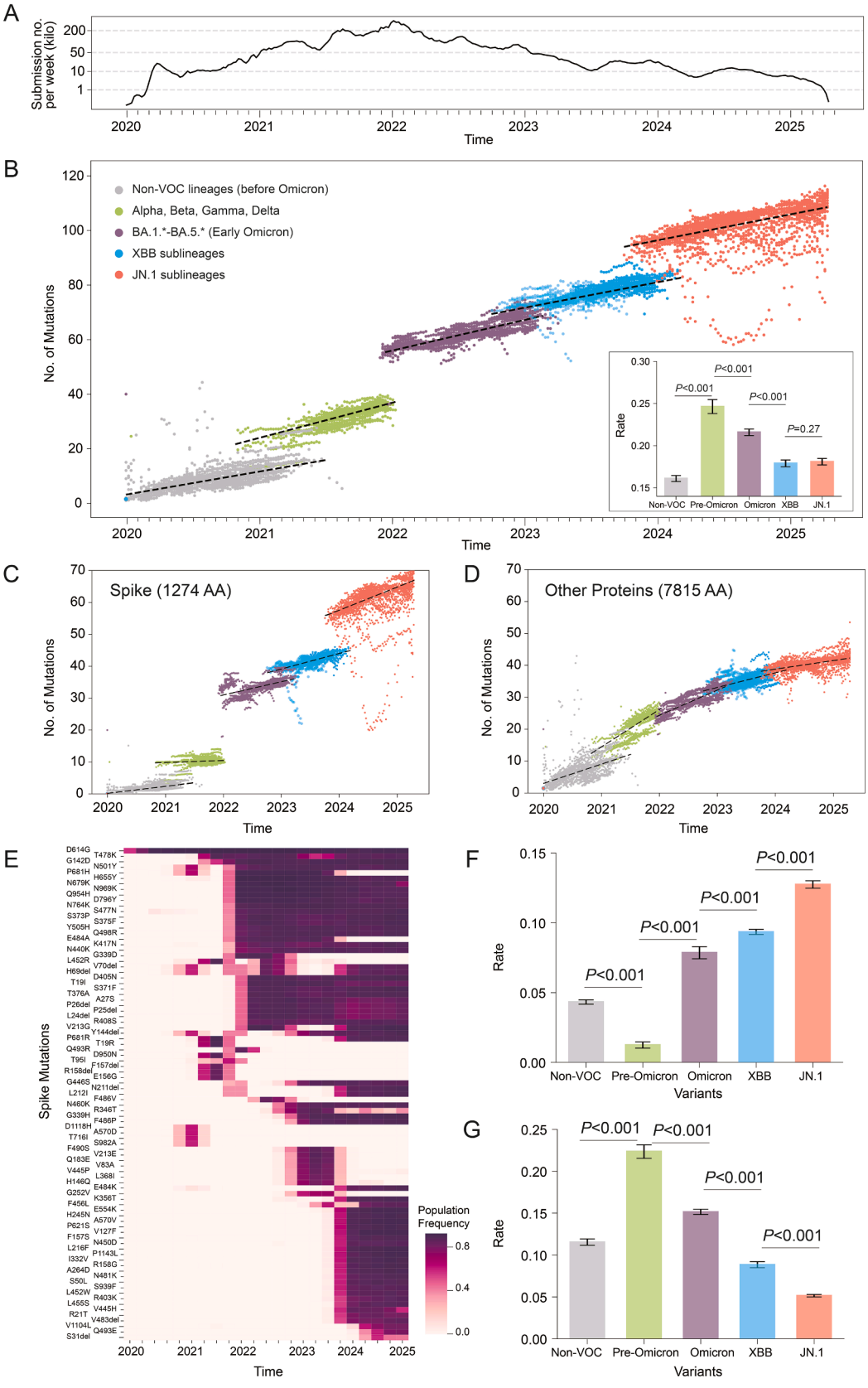
We used the PyR0 framework, a hierarchical Bayesian model developed by Obermeyer et al. (2022), to quantitatively assess fitness of viral lineages. PyR0 was originally applied to SARS-CoV-2 lineages up to Jan 2022, spanning the emergence of early Omicron lineages BA.1 and BA.2. We extended the analysis across five years. As shown in Fig. 2A, while the virus exhibited a generally continuous rise in fitness, notably, it had two times of acceleration in the rate of increasing fitness. The first acceleration coincided with the emergence of Omicron, and induced roughly a 10-fold increase in the rate of fitness increase (consistent with (Obermeyer et al., 2022)). The second significant acceleration in the rate of fitness gain was attributable to the JN.1 (approximately 1.89-fold, two-sided Wilcoxon rank-sum, $P < 0.001$ compared to the previous lineages). By contrast, the XBB did not present a significant acceleration but a slight decline in the rate (inset of Fig. 2A, see also the baseline-removed within-group analysis in Supplementary Fig. S5).

Since mutations became increasingly enriched in the S protein over time and diversification has remained high since the emergence of Omicron, we further used CoVPF, a tool we previously developed for Omicron sublineages (Lei et al., 2025), to estimate lineage fitness within Omicron. CoVPF adopts a framework similar to PyR0 while focusing on the S protein and incorporates Omicron-context functional effects (schematic diagram in Supplementary Fig. S6). As shown in Fig. 2B, the fitness increasing with time by using CoVPF is qualitatively consistent with the PyR0-based estimates. JN.1 exhibited a significantly higher acceleration rate than XBB (1.62-fold, two-sided Wilcoxon rank-sum test, $P < 0.001$) and early Omicron (1.86-fold, two-sided Wilcoxon rank-sum test, $P < 0.001$).

JN.1 had more loci harboring recurrent and shift mutations

Based on the fitness analysis, we divided the five-year evolution of SARS-CoV-2 into three stages in the following analysis: (i) the stage before the emergence of Omicron (Dec 2019 to Oct 2021, referred to as the pre-Omicron stage), (ii) the stage before the emergence of JN.1 (Nov 2021 to Oct 2023, the early-Omicron stage, including BA.1–BA.5 and XBB), and (iii) the JN.1 period (Nov 2023 to May 2025, the JN.1 stage).

Within each stage, we extracted fixed mutations, defined as specific AA changes that reached a global frequency of ≥50% in at least one 3-month window, using all genomes submitted during the corresponding period. Relative to the pre-Omicron stage, the early-Omicron stage exhibited a marked expansion of newly fixed mutations in the S protein, with 42 newly fixed over 27 months (1.55 per month). Notably, the JN.1 stage accumulated 30 newly fixed S mutations within 18 months, corresponding to an even higher rate (1.67 per month) than that observed in the early-Omicron stage (Fig. 3A). In contrast, newly fixed mutations in other proteins declined in the JN.1 stage, from 0.96 per month in the early-Omicron stage to 0.67 per month (Fig. 3B).



(caption on next page)

Fig. 1. The genomics and amino-acid-altering mutation accumulation of SARS-CoV-2 from 2020 to 2025. **A** Number of SARS-CoV-2 genomes weekly submitted to GISAID from Dec 30, 2019 to May 3, 2025. **B** Accumulation of mutations over time. Each point denotes the mean number of mutations of Pango lineages with a population frequency of no less than 0.5%. The accumulation rates by Huber regression (50% subsampling for 100 times, mean value) are indicated by the dashed lines, and distributions of fitted slopes are shown in the inset with *P* values from interaction tests. **C**, **D** Mutations in S protein (**C**) and in non-spike proteins (**D**), colored by the lineages indicated in (**B**). **E** Mutations in the S protein over time. **F**, **G** Rates of mutation accumulation via Huber regression for SARS-CoV-2 S protein (**F**) and other proteins (**G**).

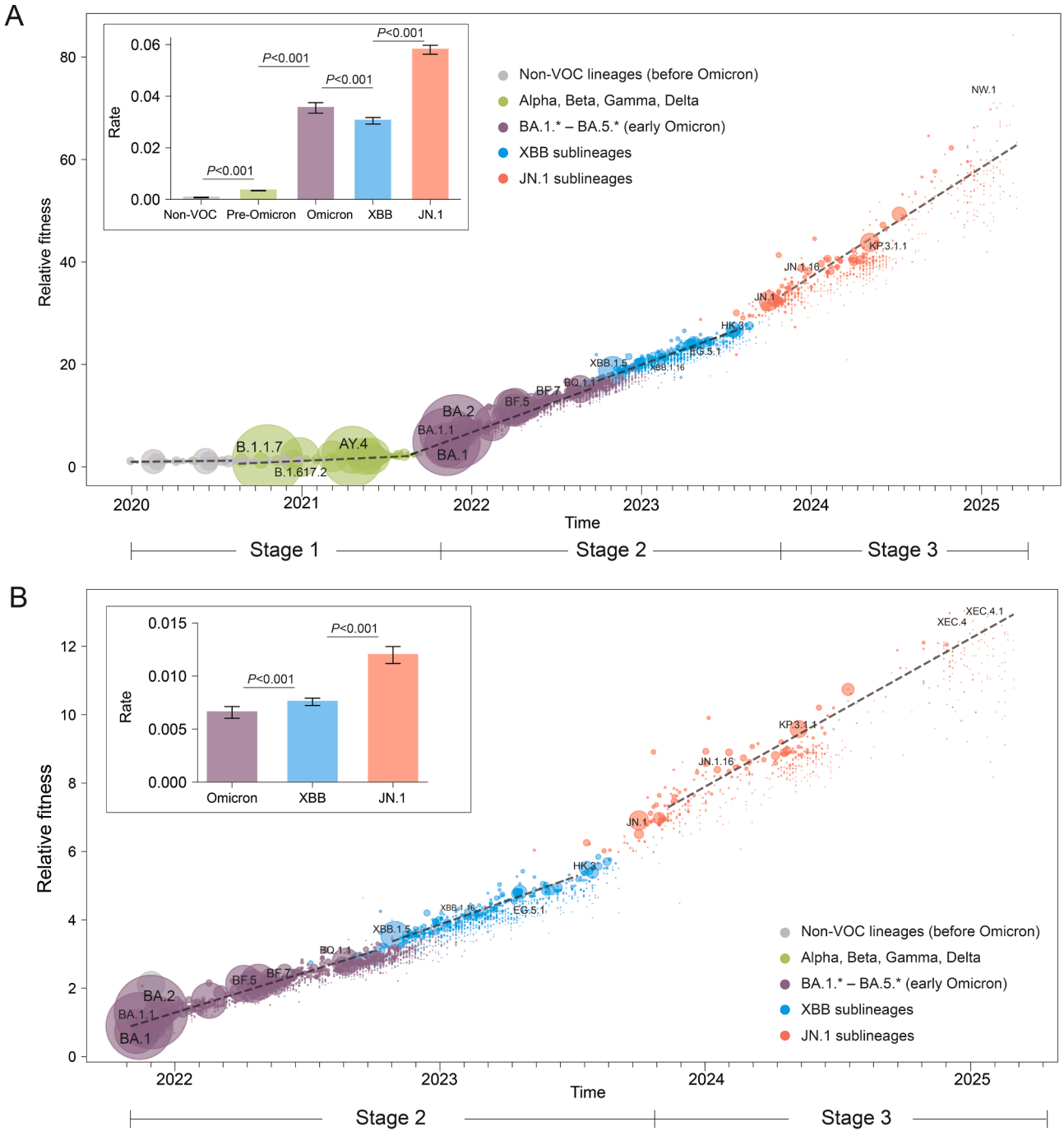


Fig. 2. Acceleration of viral fitness increasing. **A** Fitness of SARS-CoV-2 lineages estimated using PyR0. Relative fitness was calculated by subtracting the fitness of the prototypic strain estimated by PyR0 from the fitness of each lineage estimated by PyR0. The increasing rates of fitness by Huber regression are indicated by dashed lines. The inset shows the distribution of fitted regression slopes. **B** Fitness analysis of global SARS-CoV-2 lineages by using CoVPF. The relative fitness was calculated based on the Omicron ancestral strain B.1.1.529.

Notably, JN.1 stage had apparently more novel or recurrent fixed mutations at the loci that previously harbored a different fixed mutation, which we referred to as shift mutated loci (Fig. 3C). In the pre-Omicron stage, only one shift mutated locus was observed, position 681 in S protein, which successively harbored P681H in Alpha and P681R in Delta variants. The early-Omicron stage had six shift mutated loci,

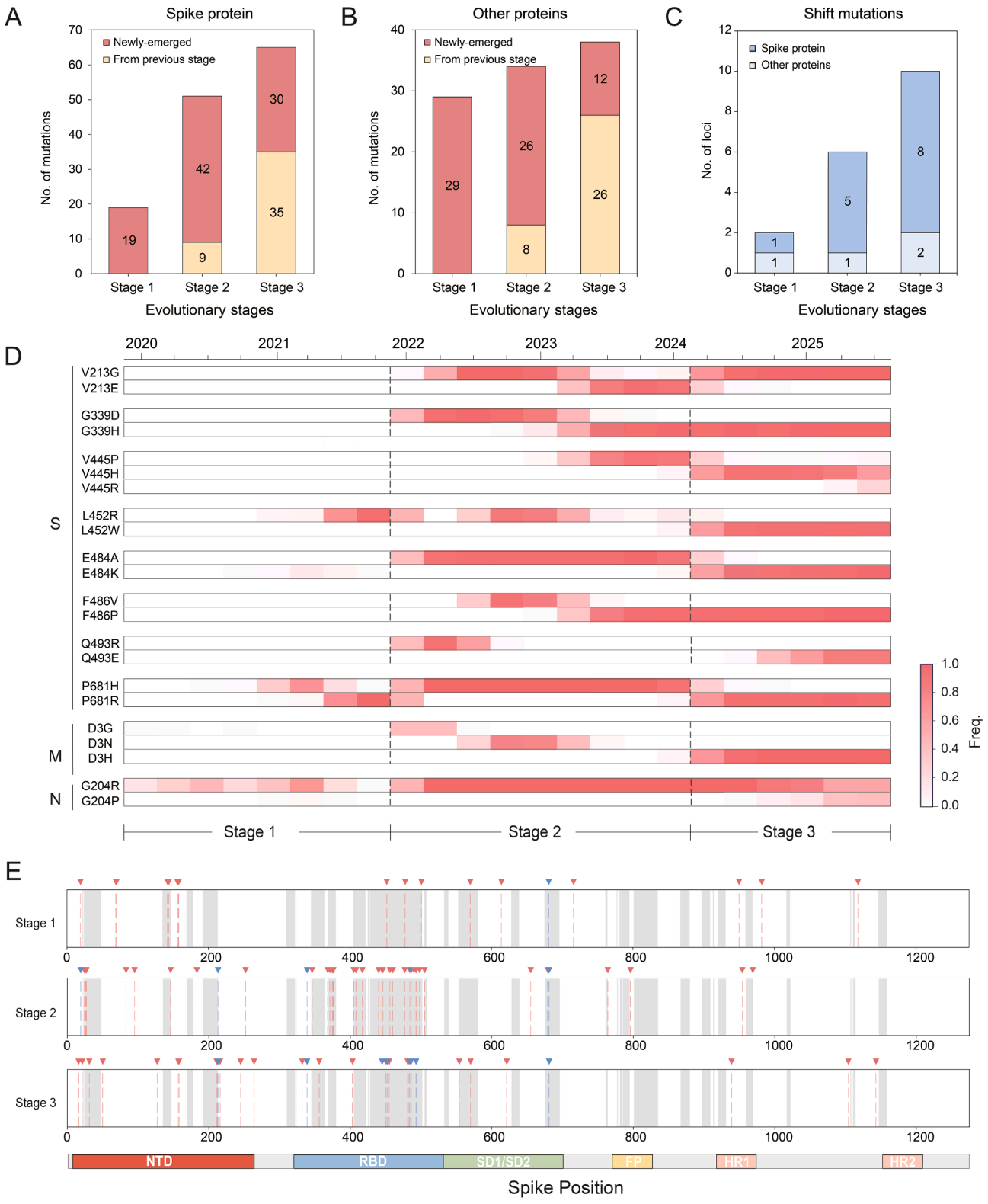


Fig. 3. Fixed amino-acid-altering mutations and loci with shift mutations. **A**, **B** Number of fixed mutations in the S protein (**A**) and the other proteins (**B**) in the stage before Omicron (stage 1), the stage of early Omicron (including XBB, stage 2) and JN.1 (stage 3). Categories indicate whether a fixed mutation was newly observed in the stage or remained fixed from the immediately preceding stage. **C** Statistics of loci with recurrent or shift mutations in each stage. **D** Recurrent or shift mutations in the JN.1 stage over time. **E** Distribution of fixed loci along the S protein across three stages and antigenic epitope regions. The structure of the S protein is shown below. Hot epitope regions are shaded in gray ($\rho = 0.014$; see Methods). The loci of shift mutations are denoted by blue triangles and blue dashed vertical lines. Other novel fixed mutations in each stage are denoted by red triangles and red dashed vertical lines.

positions 19, 213, 339, 486 and 681 in S protein, and position 3 in membrane (M) protein, in which P681H is a recurrent mutation. In the JN.1 stage, ten loci exhibited fixed shift mutations (Fig. 3D). Among these, four loci in S protein (positions 445, 452, 484 and 493), one locus in nucleocapsid (N) protein (position 204) and one locus in M protein (position 3) were first identified as shift mutated loci; while two loci in S protein carried recurrent substitutions (P681H and V213G). The six novel shift mutated loci were not inherited from the JN.1 ancestor BA.2/BA.2.86, suggesting they were more likely to be derived from adaptive evolution.

Using our curated functional annotation database AnnCovDB (Zhang et al., 2025) for a literature research, we found that seven of the eight shift mutations and twelve additional newly fixed mutations of S protein in the JN.1 stage have experimental evidence consistent with immune escape (Supplementary Table S1–2). Moreover, using hot epitope regions defined from Immune Epitope Database mapping records (see Methods), we systematically compared changes of S protein across stages. Newly fixed S mutations were significantly enriched within hot epitope regions in the early-Omicron and JN.1 stages (two-sided Fisher exact test, $P < 0.037$), but not in the pre-Omicron stage ($P = 0.448$) (Supplementary Fig. S7, Supplementary Table S3–S4). In addition, shift mutated loci showed a stronger epitope focus in JN.1, with 6/8 loci located within hot epitope regions in the JN.1 stage (two-sided Fisher exact test, $P = 0.013$) compared with 2/5 in the early-Omicron stage ($P = 0.646$) (Fig. 3E, blue dotted lines).

Mutations in JN.1 RBD had diverse physicochemical and structural features

We next examined whether JN.1 variants exhibit distinctive features at the protein structural level. Since the S protein receptor-binding domain (RBD) is enriched for fixed mutations (Fig. 3E), we extracted 225 SARS-CoV-2 RBD sequences that reached a $\geq$ 0.5% global viral population frequency, for at least one week, during 2020–2025. According to the corresponding variant lineages, these RBD sequences were assigned to groups JN.1 ($n = 88$), XBB ($n = 38$), early-Omicron ($n = 70$), pre-Omicron ($n = 13$), and non-VOC ($n = 16$) (Supplementary Table S5).

We found that those in JN.1 were more likely to induce substantial variations in both residue hydropathy and charge than other groups (Fig. 4A and B). For instance, K356T fixed in JN.1 is associated with a 78.3% increase in hydropathy and 95.5% decrease in charge; N481K and E484K shifted from neutral to positively charged states, while N450D and Q493E shifted to negatively charged. We conducted a clustering analysis based on the hydropathy and charge profiles, and the results show that the JN.1 were separated from other groups (Fig. 4C and D).

To further assess structural difference, we performed 100 ns molecular dynamics (MD) simulations for ten representative variants each from JN.1, XBB, and early-Omicron groups, together with one Alpha variant and three replicates of prototypic strain (Supplementary Table S6 and Supplementary Fig. S8 for RMSD). Based on the MD trajectories, we obtained residue root mean square fluctuation (RMSF) and Cα displacement ($|\triangle CA|$) relative to the prototypic (Fig. 4E and F). Excluding the N and C termini, the receptor binding motif (RBM, spanning residues 438 to 506) showed large fluctuations across all variants. For $|\triangle CA|$, Omicron variants exhibited consistently higher fluctuations in RBM than the Alpha and prototypic variants. However, clustering based on RMSF and $|\triangle CA|$ profiles could not separate JN.1 from other lineages (Fig. 4H and I).

We next extracted the protein structures corresponding to the lowest-energy conformations from the MD trajectories. Pairwise structural similarities (TM-scores) among these RBD variants were calculated for a subsequent multiple structural alignment (MSTA) analysis. However, by using MSTA, we could not cluster variants by their groups (Fig. 4G). It might be due to the intrinsic flexibility of the loop-rich RBD. Nonetheless, we found that solvent accessible surface area (SASA) based on the lowest-energy conformations revealed a separation of these variants by lineages, and JN.1 variants displayed lineage specific SASA profiles compared to other lineages (Fig. 4J and K).

Finally, we exhibited hydropathy, charge, and SASA profiles for representative RBD variants from JN.1, XBB, early-Omicron, and the prototypic (Fig. 4L–N). Overall, the characteristic JN.1 RBD mutations tended to increase hydrophilicity (Fig. 4L), prominently at V445H, L452W, L455S, V483del, and E484K. In terms of charge, JN.1 showed a diverse change; 5 mutations turned residues from positively charged states to neutral or negatively charged states, and 7 mutations introduced positive charges (Fig. 4M). For SASA, among RBM residues evaluated in JN.1, 5 showed decreases and 15 showed increases of more than 30% relative to the prototypic (total loci in Supplementary Fig. S9). These changes in surface exposure indicated localized large surface rearrangements within the RBD for the JN.1.

DISCUSSION

Our large-scale genomic epidemiology analysis provides a comprehensive assessment of SARS-CoV-2 fitness evolution from 2020 to 2025. We found that the emergence of Omicron coincided with a major acceleration in fitness gain and was followed by an overall linear upward trajectory in lineage fitness. Notably, JN.1 exhibited a significantly higher rate of fitness increase than early Omicron and XBB lineages, indicating a second acceleration in fitness gain. During the JN.1 stage, the S protein accumulated both newly fixed mutations and recurrent changes at sites that had already fixed different mutations previously. Furthermore, these newly fixed and shift mutations were preferentially enriched within antigenic epitope regions of S protein, consistent with intensified immune-driven selection. Physicochemical profiling and MD-based structural analyses further showed that JN.1 RBD mutations were associated with larger residual changes in physicochemical properties and surface exposure, distinguishing it from other variants. These findings indicate that JN.1 constitutes a relatively distinct evolutionary state among currently circulating Omicron lineages.

The shift mutated loci observed in the JN.1 stage, including the recurrence of P681R in S protein, are particularly noteworthy. As population immunity has accumulated, immune pressure has become a dominant selective force shaping SARS-CoV-2 evolution. Consistent with this, Cao et al. reported that convergent mutations increasingly concentrate at antibody binding sites, especially within the RBD (Cao et al., 2023). In this context, recurrent changes at previously fixed loci suggest that SARS-CoV-2 may be entering an evolutionary stage in which it repeatedly revisits and repurposes mutation loci that were established earlier, potentially expanding accessible antigenic configurations. This pattern further highlights the importance of elucidating mutational epistasis and compensatory interactions that can modulate the functional consequences of such shifts, potentially implying that mutation accumulation may not directly correspond to fitness (Moulana et al., 2022).

The fitness estimation in this study is based on genomic epidemiology and lineage frequency dynamics, which quantify realized transmission advantage as a growth-advantage estimate inferred from changes in lineage frequencies under competition in the population. These frequency-based methods, such as PyR0 (Obermeyer et al., 2022) and the approach of Li et al. (2023), infer population-level fitness from lineage frequency dynamics in genomic surveillance data, reflecting the combined effects of multiple evolutionary and epidemiological drivers. In contrast, protein- and genome-language-model approaches, such as CoVFit (Ito et al., 2025) and GIVAL (Jiang et al., 2025), learn sequence representations to predict fitness-related phenotypes or adaptation potential from sequence alone, enabling early evaluation and prioritization of newly detected or hypothetical variants. Here, we focus on frequency-based epidemiological inference to quantify population-level fitness dynamics. Frequency-based inference and sequence-based predictors are complementary and offer different perspectives on viral fitness.

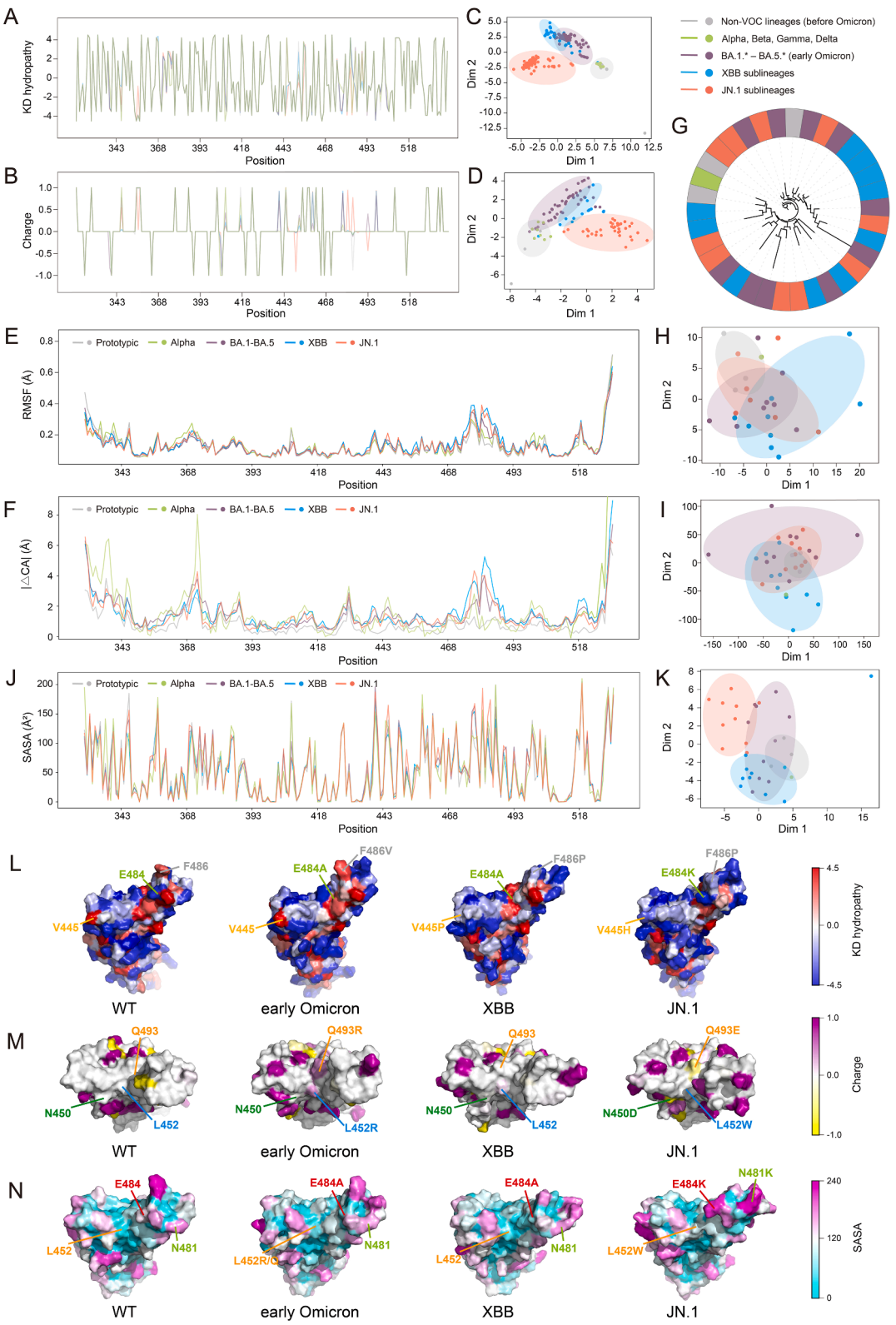


Fig. 4. Physicochemical and structural features of the RBD. **A, B** Mean residue hydropathy (**A**) and charge (**B**) profiles of viral variants from different lineages. **C, D** Two-dimensional projections from principal component analysis (PCA) of hydropathy (**C**) and charge (**D**) profiles. **E, F** Root mean square fluctuation (RMSF, **E**) and Cα displacement (|△CA|, **F**) variation profiles of the RBD relative to the prototypic strain, derived from molecular dynamics simulations. **G** Multiple structural alignment based on pairwise TM-scores among representative RBD structures. **H, I** Two-dimensional PCA projections of RMSF (**H**) and |△CA| (**I**) profiles. **J** Mean residue solvent accessible surface area (SASA) calculated from the lowest-energy conformations extracted from MD trajectories. **K** Two-dimensional PCA projection of residue SASA features. **L–N** Structural mapping of hydropathy (**L**), charge (**M**), and SASA (**N**) for representative RBD structures from the prototypic strain, early Omicron, XBB, and JN.1.

Although SARS-CoV-2 continues to circulate and evolve, genomic surveillance has declined substantially, potentially delaying the early detection of variants posing increased public health risk. According to the WHO, as of April 6, 2024, 121 countries had submitted 162,773 JN.1 genomes to GISAID (World Health Organization, 2024). By contrast, as of June 22, 2025, the newly designated VUM XFG was represented by only 1648 genomes from 38 countries, and the WHO explicitly noted low levels of sequencing data (World Health Organization, 2025). Together with our evidence for accelerated fitness gain during the rise of JN.1, these observations underscore the need for sustained and sufficiently scaled genomic surveillance to enable timely risk assessment and response.

This study has several limitations. First, our analyses were performed at the global scale and did not explicitly model regional heterogeneity. Reliable inference requires dense sampling, and sequences are disproportionately contributed by a limited number of regions, such as the United Kingdom and the United States. As a result, global estimates may be dominated by these well sampled settings and may not fully reflect evolutionary dynamics in under sampled regions. Second, our hotspot epitope analysis is based on curated database records and relies on defined thresholds, which may introduce uncertainty. Moreover, because host immunity and viral antigenicity coevolve, epitope landscapes are dynamic and may not be fully captured by static annotations. Literature research is also incomplete, as many mutations remain poorly characterized and experimental validation, particularly of epistasis interactions, is still limited. Third, although MD simulations provide a widely used approach for comparative structural analysis, they remain approximations and may not fully recapitulate experimentally determined structures such as those resolved by cryo-EM. Nevertheless, the sequence-derived physicochemical trends observed for JN.1 were consistent and remained evident under clustering analyses.

In sum, our study provides a unified, five-year assessment of SARS-CoV-2 fitness evolution and highlights the distinct evolutionary profile of JN.1. These findings underscore the importance of sustained genomic surveillance, together with timely functional and antigenic characterization, to support ongoing risk assessment as SARS-CoV-2 continues to adapt.

CONCLUSIONS

In conclusion, our analysis of 15.23 million SARS-CoV-2 genomes from December 2019 to May 2025 provides a comprehensive view of the long-term fitness evolution of SARS-CoV-2. We found that lineage fitness increased approximately linearly from 2021 to 2025, and that JN.1 marked a second acceleration in fitness gain beyond early Omicron. This transition was accompanied by continued accelerated mutation accumulation in the S protein and by more frequent recurrent and shifted mutations at previously mutated sites, as well as larger alterations in physicochemical properties and structural surface exposure within the RBD of JN.1. These findings suggest that JN.1 represents a relatively distinct evolutionary stage among currently circulating Omicron lineages and emphasize the importance of sustained genomic surveillance.

MATERIALS AND METHODS

Original genomic epidemiology data

The genomic data and related epidemiological data of SARS-CoV-2 variants were obtained from the Global Initiative on Sharing All Influenza Data (GISAID, https://gisaid.org) on May 3, 2025, including the information on sampling time, lineages by Pango nomenclature, viral variants in the metadata file, and amino acid mutations relative to hCoV-19/Wuhan/WIV04/2019 (the prototypic SARS-CoV-2 strain, GISAID accession EPI_ISL_402124, NCBI GenBank accession NC_045512). Metadata entries with complete sampling year and annotated lineage information were retained, and this resulted in a dataset comprising 15,227,368 entries for further analysis. Among these, 26 major variants were identified according to the World Health Organization (WHO) classification, based on characteristics such as transmissibility and pathogenicity. These major variants were further subdivided into 4182 lineages based on the Pango nomenclature. Since metadata with identical lineage labels were occasionally assigned to different variant groups, such ambiguities were resolved by counting the occurrence of each lineage under all variants and assigning it to the variant with the highest frequency.

Distribution of lineages and mutations

Based on the screened genomic and epidemiological dataset, we constructed a two dimensional tensor to represent lineage prevalence over time. Lineage counts were aggregated in 7-day bins from December 30, 2019 to May 3, 2025. This yielded a prevalence spectrum consisting of 277 time bins and counts for 4182 lineages globally. A lineage was considered prevalent within a given bin if it accounted for $\geq 0.5\%$ of all samples collected in that bin.

To quantify the distribution of amino acid mutations within each lineage across genes, we constructed a two dimensional tensor referred to as the mutational spectrum. In order to decrease the potential impact of low-frequency mutations on reflecting the characterization of lineages, the amino acid mutations whose frequencies were $< 20\%$ in all lineages were disregarded. The resulting mutational spectrum captured the distribution of 6012 mutations of all SARS-CoV-2 genes across 4182 lineages. Spectrum values were defined as the within-lineage frequency.

Fitness evaluation of lineages

We estimated lineage fitness using PyR0, a model proposed by Obermeyer et al. (2022). PyR0 is a hierarchical Bayesian multinomial logistic regression framework that utilizes the global spatiotemporal distribution of SARS-CoV-2 genomes and estimates lineage specific growth advantages relative to a reference lineage by variational inference. In this study, we focused on global growth rates to capture broad evolutionary trends while reducing sensitivity to geographic heterogeneity.

According to the fundamental theorem of natural selection (Crow, 2017), the variation of genotype frequencies follows a logistic curve, and the fitness of the genotypes are defined as the growth rate parameters. PyR0 models the relative frequency of lineage L at time T as:

$$P_{TL} = softmax(\alpha_L + \beta_L \times T)$$

where P_{TL} is the relative frequency of lineage L at time T, α_L represents its baseline prevalence, and β_L denotes its fitness. The SoftMax function was defined as $softmax(x_i) = exp(x_i)/sum_i^n(exp(x_i))$, where x_i equals $(\alpha + \beta \times T)$ in PyR0, i indexes a given lineage and n is the number of all lineages.

Regression analysis and growth rate comparison

To quantify temporal increases in amino acid mutation counts and relative fitness across variants, we fitted separate Huber regression models that reduce sensitivity to outliers. We performed 100 resampling iterations, each time randomly selecting 50% of observations within each time bin, and reported the mean fitted slope to obtain stable within-variant growth rate estimates. For fitness estimation, we additionally used a weighted Huber regression with weights proportional to the number of submissions, so that more representative lineages contributed more strongly to model fitting.

To assess whether growth rates differed significantly among variants, we fitted a robust linear model using Huber regression and included

interaction terms between the continuous time variable t and the variant group. The model was specified as:

$$y_t = \beta_0 + \beta_1 t + \sum_i \gamma_i \cdot variant_i + \sum_i \delta_i (t \times variant_i)$$

where y_t denotes the response at time t, defined as the mutation count or relative fitness for the reference variant, and variants are indicator variables for non-reference variants. In this parameterization, β_1 represents the growth rate of the reference variant, γ_i captures variant-specific intercept shifts relative to the reference, and δ_i is the coefficient of the time-by-variant interaction term that quantifies the difference in growth rate between $variant_i$ and the reference. We alternated the reference variant across model fits and evaluated the statistical significance of growth rate differences using the P values of the corresponding interaction terms.

Epitope annotation from Immune Epitope Database

We integrated experimentally validated epitope annotations from the Immune Epitope Database (IEDB, https://www.iedb.org) (Vita et al., 2025). Linear epitopes specific to the SARS-CoV-2 S protein were retrieved on June 12, 2025, yielding 5033 records. Each record provided the start and end amino acid positions of an epitope on S protein, derived from T cell or B cell epitope mapping studies.

To quantify epitope density along S protein, we counted, for each amino acid position (1–1273), the number of records in which the position fell within an annotated epitope interval and normalized this count by the total number of records to reduce redundancy and sampling bias. Loci with normalized epitope frequency exceeding a predefined threshold ρ were designated as hot epitope regions. We used $\rho = 0.014$ for Fig. 3, and results under alternative thresholds are shown in Supplementary Fig. S10. We evaluated multiple threshold values and assessed enrichment of mutations within hot epitope regions using a Fisher exact test (Supplementary Fig. S7 and S10).

Structural and physicochemical characterization of RBD variants

All RBD sequences that reached a weekly global population frequency $\geq 0.5\%$ between 2020 and 2025 were aligned to the prototypic reference (position 319 to 541 in S protein), and further analyses were carried out using reference numbering. Per-residue hydropathy profiles were obtained by mapping each aligned amino acid to the Kyte-Doolittle hydropathy scale, and per-residue charge profiles were defined using a simplified side-chain charge scheme at physiological pH (Asp/Glu = −1, Lys/Arg = +1, His = +0.5, all other residues = 0). For each clade, per-position mean hydropathy and charge were calculated across all sequences. These profiles were also assembled into matrices, standardized, and subjected to dimensionality reduction (PCA, UMAP, and t-SNE) for clustering analyses.

To assess structural consequences, starting from the prototypic RBD structure (PDB accession 6M0J), lineage-defining mutations of 34 representative variants were introduced using PyRosetta with side-chain repacking and local energy minimization while constraining the backbone. The resulting models were used as starting structures for 100-ns MD simulations in explicit solvent with GROMACS, using the CHARMM36 m force field and the TIP3P water model. After standard energy minimization and equilibration, 100-ns production trajectories were generated for each variant. Trajectories were aligned to the prototypic RBD using Cα atoms, and per-residue RMSF was computed from the production runs, while the lowest-energy conformation from each trajectory was extracted for subsequent analyses. Per-residue backbone displacements $|\triangle CA|$ were calculated as the Euclidean distance between corresponding Cα atoms. SASA was computed on the lowest-energy conformations using the Shrake-Rupley algorithm with a 1.4 Å probe, and per-residue SASA was obtained by summing over atoms in each residue.

For visualization, clade-wise mean hydropathy, charge, and SASA values were mapped to representative RBD structures from the prototypic, early-Omicron, XBB, and JN.1 groups, and colored in PyMOL to depict position-specific physicochemical changes on the RBD surface.

DATA AVAILABILITY

All data analyzed in this study were obtained from the Global Initiative on Sharing All Influenza Data (GISAID, https://gisaid.org) database.

ETHICS STATEMENT

This article does not contain any studies with human or animal subjects performed by any of the authors.

AUTHOR CONTRIBUTIONS

Zhong-yi Lei: Conceptualization, Methodology, Validation, Formal analysis, Writing-Original Draft, Visualization. Qian-tong Jin: Methodology, Software, Validation. Xiao-min Zhang: Investigation, Data Curation. Xiao-chen Bo: Resources, Supervision. Zi-lin Ren: Methodology, Supervision. Yi-gang Tong: Supervision, Funding acquisition. Ming Ni: Conceptualization, Writing-Review & Editing, Supervision, Project administration.

CONFLICT OF INTEREST

The authors declare that there are no conflicts of interest regarding the publication of this paper. Prof. Yi-gang Tong is editorial board member for *Virologica Sinica* and was not involved in the editorial review or the decision to publish this article.

ACKNOWLEDGEMENTS

This work is supported by the National Natural Science Foundation of China (Grant nos. 92369201), and the Capital's Funds for Health Improvement and Research (Grant nos. 2024-1G-4421).

APPENDIX A. SUPPLEMENTARY DATA

Supplementary data to this article can be found online at https://doi.org/10.1016/j.virs.2026.03.015.

REFERENCES

Cao, Y., Jian, F., Wang, J., Yu, Y., Song, W., Yisimayi, A., Wang, J., An, R., Chen, X., Zhang, N., Wang, Yao, Wang, P., Zhao, L., Sun, H., Yu, L., Yang, S., Niu, X., Xiao, T., Gu, Q., Shao, F., Hao, X., Xu, Y., Jin, R., Shen, Z., Wang, Youchun, Xie, X.S., 2023. Imprinted SARS-CoV-2 humoral immunity induces convergent Omicron RBD evolution. Nature 614, 521–529.

Chaguza, C., Coppi, A., Earnest, R., Ferguson, D., Kerantzas, N., Warner, F., Young, H.P., Breban, M.I., Billig, K., Koch, R.T., Pham, K., Kalinich, C.C., Ott, I.M., Fauver, J.R., Hahn, A.M., Tikhonova, I.R., Castaldi, C., De Kumar, B., Pettker, C.M., Warren, J.L., Weinberger, D.M., Landry, M.L., Peaper, D.R., Schulz, W., Vogels, C.B.F., Grubaugh, N.D., 2022. Rapid emergence of SARS-CoV-2 Omicron variant is associated with an infection advantage over Delta in vaccinated persons. Med 3, 325–334.e4.

Crow, J.F., 2017. An Introduction to Population Genetics Theory. Scientific Publishers.

Guo, C., Yu, Y., Liu, J., Jian, F., Yang, S., Song, W., Yu, L., Shao, F., Cao, Y., 2025. Antigenic and virological characteristics of SARS-CoV-2 variants BA.3.2, XFG, and NB.1.8.1. Lancet Infect. Dis. 25, e374–e377.

Harvey, W.T., Carabelli, A.M., Jackson, B., Gupta, R.K., Thomson, E.C., Harrison, E.M., Ludden, C., Reeve, R., Rambaut, A., Peacock, S.J., Robertson, D.L., 2021. SARS-CoV-2 variants, spike mutations and immune escape. Nat. Rev. Microbiol. 19, 409–424.

Ioannou, G.N., Berry, K., Yan, L., Huang, Y., Lin, H.-M., Bui, D., Hynes, D.M., Boyko, E.J., Ferguson, J.M., Aslan, M., Bajema, K.L., 2025. Effectiveness of the 2024–2025 KP.2 COVID-19 vaccines in the United States during long-term follow-up. Nat. Commun. 17, 1043.

Ito, J., Strange, A., Liu, W., Joas, G., Lytras, S., Sato, K., 2025. A protein language model for exploring viral fitness landscapes. Nat. Commun. 16, 4236.

Jian, F., Wang, J., Yisimayi, A., Song, W., Xu, Y., Chen, X., Niu, X., Yang, S., Yu, Y., Wang, P., Sun, H., Yu, L., Wang, J., Wang, Yao, An, R., Wang, W., Ma, M., Xiao, T., Gu, Q., Shao, F., Wang, Youchun, Shen, Z., Jin, R., Cao, Y., 2025. Evolving antibody response to SARS-CoV-2 antigenic shift from XBB to JN.1. Nature 637, 921–929.

Jiang, S.-Y., Zhao, S.-S., Wei, J.-Q., Zhang, S., Zhao, Z., Tong, Y., Liu, W., Wang, J., Jiang, T., Li, J., 2025. General intelligence framework to predict virus adaptation based on a genome language model. Research 8, 871.

Lei, Z., Zhang, X., Han, J., Xue, J., Xu, J., Ren, Z., Tong, Y., Bo, X., Ni, M., 2025. Integrating genomic epidemiology and deep mutational scanning data for prevalence forecasting of SARS-CoV-2 Omicron lineages. PLoS One 20, e0335520.

Li, J., Lai, S., Gao, G.F., Shi, W., 2021. The emergence, genomic diversity and global spread of SARS-CoV-2. Nature 600, 408–418.

Li, X., Yan, H., Wong, G., Ouyang, W., Cui, J., 2023. Identifying featured indels associated with SARS-CoV-2 fitness. Microbiol. Spectr. 11, e02269-23.

Li, J., Yang, J., Ding, X., Zhou, H., Han, N., Wu, A., 2024. The spatiotemporal analysis of SARS-CoV-2 transmission in China since the termination of the dynamic zero-COVID policy. Virol. Sin. 39, 737–746.

Liu, J., Yu, Y., Yang, S., Jian, F., Song, W., Yu, L., Shao, F., Cao, Y., 2025. Virological and antigenic characteristics of SARS-CoV-2 variants LF.7.2.1, NP.1, and LP.8.1. Lancet Infect. Dis. 25, e128–e130.

Ma, W., Fu, H., Jian, F., Cao, Y., Li, M., 2023. Immune evasion and ACE2 binding affinity contribute to SARS-CoV-2 evolution. Nat. Ecol. Evol. 7, 1457–1466.

Markov, P.V., Ghafari, M., Beer, M., Lythgoe, K., Simmonds, P., Stilianakis, N.I., Katzourakis, A., 2023. The evolution of SARS-CoV-2. Nat. Rev. Microbiol. 21, 361–379.

Moulana, A., Dupic, T., Phillips, A.M., Chang, J., Nieves, S., Roffler, A.A., Greaney, A.J., Starr, T.N., Bloom, J.D., Desai, M.M., 2022. Compensatory epistasis maintains ACE2 affinity in SARS-CoV-2 Omicron BA.1. Nat. Commun. 13, 7011.

Obermeyer, F., Jankowiak, M., Barkas, N., Schaffner, S.F., Pyle, J.D., Yurkovetskiy, L., Bosso, M., Park, D.J., Babadi, M., MacInnis, B.L., Luban, J., Sabeti, P.C., Lemieux, J. E., 2022. Analysis of 6.4 million SARS-CoV-2 genomes identifies mutations associated with fitness. Science 376, 1327–1332.

Paton, R.S., Overton, C.E., Ward, T., 2022. The rapid replacement of the SARS-CoV-2 Delta variant by Omicron (B.1.1.529) in England. Sci. Transl. Med. 14, eabo5395.

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.

Viana, R., Moyo, S., Amoako, D.G., Tegally, H., Scheepers, C., Althaus, C.L., Anyaneji, U. J., Bester, P.A., Boni, M.F., Chand, M., Choga, W.T., Colquhoun, R., Davids, M., Deforche, K., Doolabh, D., du Plessis, L., Engelbrecht, S., Everatt, J., Giandhari, J., Giovanetti, M., Hardie, D., Hill, V., Hsiao, N.-Y., Iranzadeh, A., Ismail, A., Joseph, C., Joseph, R., Koopile, L., Kosakovsky Pond, S.L., Kraemer, M.U.G., Kuate-Lere, L., Laguda-Akingba, O., Lesetedi-Mafoko, O., Lessells, R.J., Lockman, S., Lucaci, A.G., Maharaj, A., Mahlangu, B., Maponga, T., Mahlakwane, K., Makatini, Z., Marais, G., Maruapula, D., Masupu, K., Matshaba, M., Mayaphi, S., Mbhele, N., Mbulawa, M.B., Mendes, A., Mlisana, K., Mnguni, A., Mohale, T., Moir, M., Moruisi, K., Mosepele, M., Motsatsi, G., Motswaledi, M.S., Mphoyakgosi, T., Msomi, N., Mwangi, P.N., Naidoo, Y., Ntuli, N., Nyaga, M., Olubayo, L., Pillay, S., Radibe, B., Ramphal, Y., Ramphal, U., San, J.E., Scott, L., Shapiro, R., Singh, L., Smith-Lawrence, P., Stevens, W., Strydom, A., Subramoney, K., Tebeila, N., Tshiabuila, D., Tsui, J., van Wyk, S., Weaver, S., Wibmer, C.K., Wilkinson, E., Wolter, N., Zarebski, A.E., Zuze, B., Goedhals, D., Preiser, W., Treurnicht, F., Venter, M., Williamson, C., Pybus, O.G., Bhiman, J., Glass, A., Martin, D.P., Rambaut, A., Gaseitsiwe, S., von Gottberg, A., de Oliveira, T., 2022. Rapid epidemic expansion of the SARS-CoV-2 Omicron variant in Southern Africa. Nature 603, 679–686.

Vita, R., Blazeska, N., Marrama, D., , IEDB Curation Team Members, Duesing, S., Bennett, J., Greenbaum, J., De Almeida Mendes, M., Mahita, J., Wheeler, D.K., Cantrell, J.R., Overton, J.A., Natale, D.A., Sette, A., Peters, B., 2025. The immune epitope database (IEDB): 2024 update. Nucleic Acids Res. 53, D436–D443.

Wang, Q., Mellis, I.A., Ho, J., Bowen, A., Kowalski-Dobson, T., Valdez, R., Katsamba, P. S., Wu, M., Lee, C., Shapiro, L., Gordon, A., Guo, Y., Ho, D.D., Liu, L., 2024. Recurrent SARS-CoV-2 Spike Mutations Confer Growth Advantages to Select JN.1 Sublineages, vol. 13. Emerging Microbes & Infections, 2402880.

World Health Organization, 2024. Updated Risk Evaluation of JN, vol. 1. https://www.who.int/docs/default-source/coronaviruse/15042024_jn1_ure.pdf.

World Health Organization, 2025 WHO TAG-VE Risk Evaluation for SARS-CoV-2 Variant Under Monitoring: XFG. https://www.who.int/docs/default-source/coronaviruse/25062025_xfg_ire.pdf.

Xu, K., An, Y., Liu, X., Xie, H., Li, D., Yang, T., Duan, M., Wang, Y., Zhao, X., Dai, L., Gao, G.F., 2024. Neutralization of SARS-CoV-2 KP.1, KP.1.1, KP.2 and KP.3 by human and murine sera. npj Vaccines 9, 215.

Yajima, H, Nomai, T, Okumura, K, Maenaka, K, Ito, J, Hashiguchi, T, Sato, K, 2024. Molecular and structural insights into SARS-CoV-2 evolution: from BA.2 to XBB subvariants. mBio 15, e03220–23.

Yang, S., Yu, Y., Xu, Y., Jian, F., Song, W., Yisimayi, A., Wang, P., Wang, J., Liu, J., Yu, L., Niu, X., Wang, J., Wang, Yao, Shao, F., Jin, R., Wang, Youchun, Cao, Y., 2024. Fast evolution of SARS-CoV-2 BA.2.86 to JN.1 under heavy immune pressure. Lancet Infect. Dis. 24, e70–e72.

Zhang, X., Lei, Z., Zhang, J., Yang, T., Liu, X., Xue, J., Ni, M., 2025. AnnCovDB: a manually curated annotation database for mutations in SARS-CoV-2 spike protein. Database 2025, baaf002.