

Research Article

Dynamic expedition of leading mutations in SARS-CoV-2 spike glycoproteins

Muhammad Hasan^{a,b,1}, Zhouyi He^{a,b,1}, Mengqi Jia^{a,1}, Alvin C.F. Leung^{a,c,1},
Kathiresan Natarajan^{d,1}, Wentao Xu^a, Shanqi Yap^a, Feng Zhou^a, Shihong Chen^a, Hailei Su^{e,*},
Kaicheng Zhu^{a,*}, Haibin Su^{a,b,*}

^a Department of Chemistry, The Hong Kong University of Science and Technology, Clear Water Bay, Hong Kong, China

^b Southern Marine Science and Engineering Guangdong Laboratory (Guangzhou), Guangzhou, China

^c Division of Life Science, The Hong Kong University of Science and Technology, Clear Water Bay, Hong Kong, China

^d Rajiv Gandhi Centre for Biotechnology, Thiruvananthapuram 695014, Kerala, India

^e Bengbu Hospital of Traditional Chinese Medicine, 4339 Huai-shang Road, Anhui 233080, China



ARTICLE INFO

Dataset link: <https://github.com/hbsulab/L-index.git>

Dataset link: <https://hbsulab.github.io/deLemus/>

ABSTRACT

The continuous evolution of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), which caused the recent pandemic, has generated countless new variants with varying fitness. Mutations of the spike glycoprotein play a particularly vital role in shaping its evolutionary trajectory, as they have the capability to alter its infectivity and antigenicity. We present a time-resolved statistical method, Dynamic Expedition of Leading Mutations (deLemus), to analyze the evolutionary dynamics of the SARS-CoV-2 spike glycoprotein. The proposed *L*-index of the deLemus method is effective in quantifying the mutation strength of each amino acid site and outlining evolutionarily significant sites, allowing the comprehensive characterization of the evolutionary mutation pattern of the spike glycoprotein.

1. Introduction

The widespread circulation of coronavirus disease 2019 (COVID-19) since its initial outbreak in December 2019, which led to a pandemic, has brought tremendous impacts on global health and the world economy [1,2]. While society has gradually returned to normalcy since the end of the COVID-19 public health emergency declared by the World Health Organization (WHO) on May 5, 2023 [3], the evolution of its etiological agent, severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), continues unabated. Unlike other RNA viruses, coronaviruses encode a 3'-to-5' exoribonuclease that binds to their low-fidelity RNA-dependent RNA polymerase for proofreading RNA replication [4–7]. However, this proofreading mechanism is not flawless, as the viral genome is still prone to substitutions, indels, and recombinations [8–11], from which thousands of SARS-CoV-2 lineages have been generated [12]. Within the first two years of the pandemic, the WHO designated five variants of concern (VOCs): Alpha (α , B.1.1.7), Beta (β , B.1.351), Gamma (γ , P.1), Delta (δ , B.1.617.2), Omicron (\omicron , B.1.1.529);

and numerous variants of interest (VOIs) [13]. In rapid succession, variants emerged, proliferated, and outcompeted their antecedents. Each wave of resurgence introduced many novel mutations that continue to finetune the fitness of the virus [10]. While beneficial mutations are generally rare [14–18], constant genetic diversification of SARS-CoV-2 has significantly hindered the development of COVID-19 treatments and infection control strategies [19].

Across the mutational landscape of the SARS-CoV-2 genome, the spike gene is one region that sits atop its plateau [20–22]. The spike glycoprotein is a trimeric type I viral fusion protein that binds the SARS-CoV-2 virion to the angiotensin-converting enzyme 2 (ACE2) receptor of a host cell [23]. Each monomer is composed of the S1 and S2 subunits that contain four functional domains: the S1 facilitates ACE2-binding [24,25], and contains the N-terminal domain (NTD), receptor-binding domain (RBD), and two C-terminal domains (CTDs); while the S2 mediates membrane fusion [26]. Because of its crucial role in cell entry [23], the spike is a primary target for antibodies [27,28], immune effector cells [29,30], and COVID-19 therapeutics [31–33]. All

* Corresponding authors.

E-mail addresses: shl75717@126.com (H.L. Su), kczechu@ust.hk (K.C. Zhu), haibinsu@ust.hk (H.B. Su).

¹ These authors contribute equally.

<https://doi.org/10.1016/j.csbj.2024.05.037>

Received 10 February 2024; Received in revised form 21 May 2024; Accepted 21 May 2024

Available online 24 May 2024

2001-0370/© 2024 The Author(s). Published by Elsevier B.V. on behalf of Research Network of Computational and Structural Biotechnology. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

these selective pressures contribute to its high mutational activity, with the spike mutations capable of altering its infectivity and antigenicity [10,11,18,19,34]. Nevertheless, the evolutionary trajectory of the spike is intricate. Out of its complete sequence space spanning over 20^{1000} unique spike sequences, only a few million are documented in the Global Initiative on Sharing Avian Influenza Data (GISAID) database [35]. Furthermore, while data shared via GISAID show that nearly all spike residues have undergone at least one mutation since the COVID-19 outbreak, only a few sites exhibit polymorphism. Although spike evolution is a highly dynamic process, there exist many constraints that prevent it from navigating through most sequence options [36–39]. Having a strategy to distinguish evolutionarily significant sites from less important ones would be a stepping stone to make sense of the logic behind SARS-CoV-2 evolution.

This work presents a time-resolved statistical method named Dynamic Expedition of Leading Mutations (deLemus) to quantitatively characterize the robust properties of SARS-CoV-2 spike glycoprotein evolution. Our proposed deLemus method combines the degree of single amino acid polymorphism with the frequency of mutations in the sequence space to generate a monthly value at each amino acid site of the spike, called the L -index, which serves as a way to outline sites that have the potential to exhibit evolutionary significance. The degree of single amino acid polymorphism encapsulates the number of residue options that a site can adopt in each month, and it carries information regarding the adaptability of a particular sequence within the viral quasispecies [37,40,41]. Indeed, theoretical and experimental studies have shown that temporal variations in polymorphisms and genetic diversities can imply features of evolutionary trade-offs under selective pressures and viral phenotypes [42–45]. The frequency of mutations in the sequence space denotes the number of non-degenerate sequences bearing certain mutations at a given site in each month, which can be used to represent the site-specific mutation strengths of a sample of strictly unique viral sequences. The analysis of the evolutionary trajectory of the SARS-CoV-2 spike glycoprotein using the L -index demonstrates its effectiveness in outlining evolutionarily significant sites.

2. Methodology

Detecting dynamic patterns from big data sets has always been a major challenge in data analysis. In this work, we propose a method called deLemus to investigate the evolutionary dynamics of the SARS-CoV-2 spike glycoprotein at an amino acid sequence level.

By the end of 2023, we had downloaded more than 15 million SARS-CoV-2 spike glycoprotein amino acid sequences from the GISAID hCoV-19 database [35]. We used EPI_ISL_402124 as the reference sequence of the spike glycoprotein [1,46]. Since there are a substantial number of repeated sequences in the original data, all degenerate sequences were removed before further analysis of sequence mutations. Overall, 667,213 non-degenerate sequences were retrieved from the entire set of reported sequences uploaded to GISAID between January 2020 and December 2023 (Fig. S1).

Sequences submitted within the same month were grouped together. Multiple sequence alignment was then consecutively conducted on each group using Clustal Omega to check the occurrence of substitution or deletion at each amino acid site [47], relative to the reference sequence. This yielded the total number of mutated sites in all sequences n and the number of sequences $P(n)$ with a given n in each month. The mutation rate Ξ in the unit $\text{seq}^{-1}\text{mo}^{-1}$ was calculated based on the total number of mutations per sequence per month (Fig. S2). The total number of single amino acid polymorphisms at each j^{th} amino acid site in each t^{th} month $s_j(t)$ and the number of amino acid sites $N(s)$ were also calculated, from which a Poisson distribution was observed, giving the monthly average of amino acid polymorphism number \bar{s} (Fig. S3).

For each t^{th} month, one $m \times l$ mutation matrix $H(t)$ was constructed based on the multiple sequence alignment data, where m is the number of non-degenerate sequences displayed in a particular month, and l

is the length of the spike glycoprotein amino acid sequence. In other words, each row represents one non-degenerate sequence from that month, and each column corresponds to one residue in the sequence. For the i^{th} sequence, if the j^{th} residue is changed, the corresponding matrix component $H_{ij}(t)$ would be set to 1. Otherwise, it would be set to 0.

We then factorized $H(t)$ by singular value decomposition [48],

$$H(t) = P(t) \cdot \Sigma(t) \cdot N(t), \quad (1)$$

the $P(t)$ is an $m \times m$ matrix and $N(t)$ is an $l \times l$ matrix containing the eigenvector information of the sequence and site, respectively, where eigenvalues are recorded at $\Sigma(t)$. (See Supplemental Material for details.)

From the monthly $\Sigma_i(t)$ and $N_{ij}(t)$, we collected the top four leading sets of mutations to compute the L -index $L_j(t)$,

$$L_j(t) = s_j(t) \cdot \sqrt{\sum_{i=1}^4 N_{ij}^2(t)}, \quad (2)$$

which is used to quantify the mutation strength of each j^{th} site in each t^{th} month. The evolutionary behavior of each leading set and their corresponding contribution to the spike evolution is visually depicted in Fig. S4. The amino acid sites were ranked according to their $L_j(t)$, and the top-ranked ones were identified as the leading mutations of each month from January 2020 to December 2023.

3. Results and discussion

The structural information for each confirmed leading mutation site was determined using AlphaFold2 [49], as illustrated in Fig. 1. In the figure, we highlighted the mutation sites observed in the reported variants. To facilitate a comprehensive analysis, we grouped the identified leading mutations into four distinct protein domains, namely NTD, RBD, CTDs, and S2. Within each domain, approximately ten leading mutations were selected every month. In addition, mutations of each VOC/VOI characterized by the WHO were compiled to give a list of ‘reported mutations’. We conducted a comparative analysis between the reported mutations and our outlined leading mutations, noting any mutations that match as ‘confirmed mutations’.

3.1. N-terminal domain (NTD)

The NTD is an S1 ectodomain located at the outermost region of the SARS-CoV-2 spike glycoprotein, where several epitopes lie [18,50–52]. While the NTD does not directly interact with ACE2 receptors, the domain’s close spatial proximity to the RBD has enabled some of its mutations to alter the cell entry dynamics of SARS-CoV-2 [53,54]. In the NTD, insertions and deletions have been frequently reported, indicating the structural plasticity within NTD [18,55]. Deletions in this domain have been characterized to frequently occupy particular locations of the spike gene known as recurrent deletion regions (RDRs), where partial nucleotide removals within specific stretches of codons can give rise to new nucleotide arrangements [8,56]. Given their substantial influence on viral antigenicity and infectivity, it is of utmost importance to closely monitor the evolutionary trajectory of the NTD.

Our investigation began with the L -index calculation of each NTD amino acid site, starting from 2020, to outline potential leading mutations. In January 2020, we identified ten mutations within the NTD (Fig. 2). Most of the outlined mutations are deletions, $\Delta I68$ -G75, which have been identified as RDR1 [8]. The $\Delta H69$ /V70 double deletion is persistently outlined until the end of 2020 and eventually appears in the Alpha NTD, contributing to the variant fitness by enhancing infectivity via syncytia formation [57]. In February 2020, we detected another deletion region ($\Delta L141$ -V143) within the NTD. These deletions were located within RDR2, which occupies the N2 loop of the

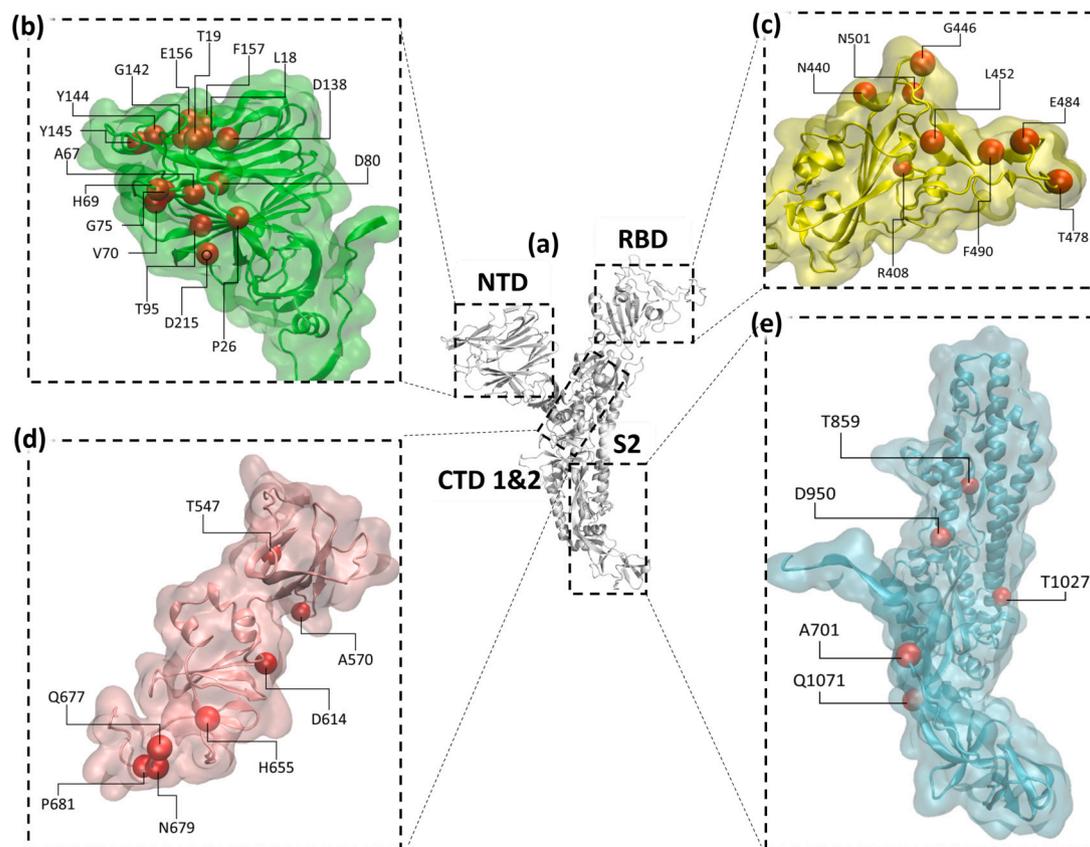


Fig. 1. Confirmed leading mutation sites (red spheres) in different domains of the SARS-CoV-2 spike glycoprotein. (a) A SARS-CoV-2 spike glycoprotein monomer (generated by AlphaFold2, Colab version, using the reference sequence EPI_ISL_402124) encompassing the four major functional domains: NTD, RBD, CTDs (CTD1&2), and S2. (b) Mutation sites outlined in the NTD (green), most of which are clustered in the outer regions. (c) Mutation sites outlined in the RBD (yellow), nearly all the captured mutations are located in the receptor-binding motif. (d) Mutation sites outlined in the CTDs (pink). (e) Mutation sites outlined in S2 (cyan).

NTD [8,58,59]. These deletions were found in the Omicron sublineage, BA.1(Δ G142/V143), and were designated as a VOC in late 2021 [2].

Similarly, we had outlined another deletion region (Δ L242-A243), which has been reported in RDR4 of the N5 loop in another study [8]. This deletion, present in the Beta variant, designated as a VOC in December 2020, contributed to the restructuring of the N5 loop, enabling evasion from monoclonal antibody 4A8 [60]. In the N3 loop, we outlined another deletion (Δ Y144) that has been shown to impair the neutralizing activities of multiple antibodies [58,59]. This deletion was found in the Alpha variant that was designated as a VOC in late 2020 [2]. In this work, we could see the emergence of multiple deletion regions (Fig. S5), providing time-resolved information on their formation, which was absent in previous studies [8].

In addition to deletions, our analysis between April and December 2020 revealed the presence of substitutions at multiple leading sites. (Fig. 2). We outlined several substitutions, including L5F, L18F, P26S, G75V, T76I, D138Y, and D253G. In particular, the L18F mutation, found in the Beta and Gamma NTDs, has been associated with immune escape capabilities [59,60]. Also, the two leading mutations of the Gamma variant (P26S and D138Y) have been found to disrupt the epitope targeted by monoclonal antibody (mAb) 159, resulting in a significant reduction in its neutralizing activity [61]. We additionally identified a mutable site, D215, exhibiting multiple amino acid polymorphisms (A, G, H, Y), which has been associated with immune escape capabilities and observed in the Beta NTD [59,60].

Between January and May 2021, we identified multiple potential leading mutations in the NTD (Fig. 2), including the Δ E156/F157 double deletion, which has been found to enhance viral fitness by evading NTD-targeting antibodies and is present in the Delta variant [62]. Ad-

ditionally, we observed polymorphism G142D within RDR2. While the effects of G142 mutations are understudied, their location within the NTD antigenic supersite suggests potential alterations in spike-antibody interactions [59]. Notably, the G142D mutation has been shown to confer significant resistance against NTD-targeting monoclonal antibodies [63]. This mutation, together with A27S, is also found in the BA.2 variant, a prominent sublineage of the Omicron variant that emerged in November 2021. Overall, the dominance of deletion and substitution mutations in the NTD indicates its dynamic evolutionary nature. Our findings underscore the importance of studying the emergence of deletion regions and mutational polymorphism, providing valuable insights for vaccine and drug development strategies to effectively counter virus evolution.

3.2. Receptor-binding domain (RBD)

The RBD located in the S1 domain, not only plays an essential role in ACE2 recognition [64,65], but also acts as a region of immunodominance targeted by around 90% of all plasma or serum neutralizing antibodies [18,66,67]. Mutations in this domain therefore often have the ability to alter virus-ACE2 or virus-antibody binding affinities [68], enabling the generation of variants with higher transmissibility or immune escape capabilities [18,69]. In fact, as shown in Fig. 1, most mutations in this domain are located in the receptor-binding motif (RBM, 438-506) that serves as the spike-ACE2 binding interface. These mutations would therefore potentially affect the infectivity of the virus. With such significant functional implications, it is necessary to track RBD mutations over time.



Fig. 2. Evolutionary trajectory of the N-terminal domain (NTD). The x-axis denotes mutable sites within the domain. The y-axis represents time in months, spanning from January 2020 to December 2021. The Variants of Concern (VOCs) are represented by the following colors: red (α , B.1.1.7), light blue (β , B.1.351), purple (γ , P.1), green (δ , B.1.617.2), and blue (ϵ , BA.1, BA.2). The Variants of Interest (VOIs) are depicted in pink: ϵ (B.1.427, B.1.429), ι (B.1.526), η (B.1.525), κ (B.1.617.1), λ (C.37), and μ (B.1.621). Reported mutations from VOCs and VOIs are depicted in gray. The orange circles represent the mutation outlined by the L -index. The vertical dashed lines indicate the confirmed mutations identified by the L -index within the reported VOCs/VOIs. Each circle is labeled with a letter representing the corresponding mutated amino acid, such as Δ (Deletion), A (Alanine), C (Cysteine), D (Aspartic Acid), etc. Unmarked circles correspond to the same amino acids as the labeled circles below.

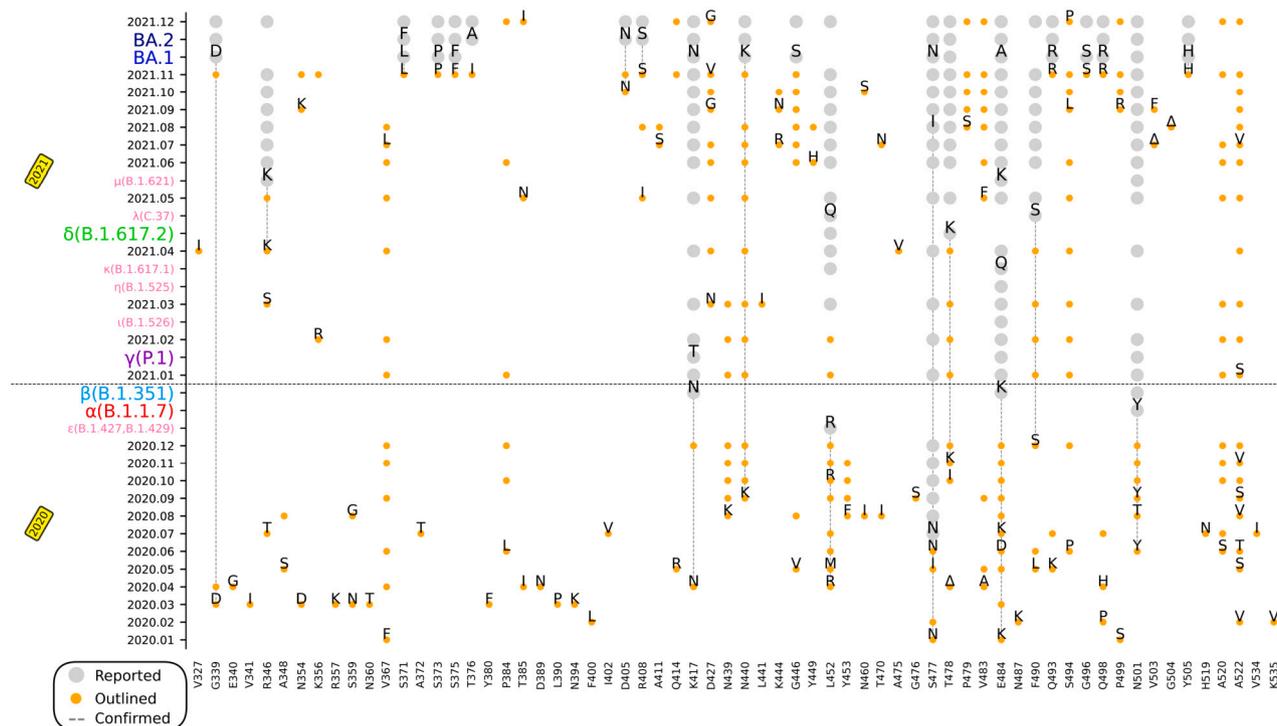


Fig. 3. Evolutionary trajectory of the receptor-binding domain (RBD). The annotations, labels, and color schemes in this figure are identical to those employed in Fig. 2.

Since the emergence of the first two reported variants, different SARS-CoV-2 variants have continued to acquire mutations in the RBD. Several of these RBD mutations were successfully outlined by our deLemus analysis. One such important mutation is N501Y, which appeared as a leading mutation in June 2020 in our deLemus analysis. This mutation can be found in the Alpha variant and the subsequent Beta, Gamma, Mu (μ , B.1.621.1), and Omicron variants, as depicted in Fig. 3. Studies have shown that the N501Y mutation can enhance ACE2 binding affinity by introducing π - π interactions between N501Y of RBD and Y41 of ACE2 [61]. E484K is another notable mutation in the Beta variant that was outlined as a leading mutation from January 2020 by our deLemus analysis (Fig. 3). Mutations at this site have been demonstrated to significantly reduce the neutralization titers of convalescent plasma [71–73]. Although some studies have suggested that mutations of the E484 residue would lead to diminished electrostatic complementarity between the RBM and the ACE2 receptor [74], many structural biology studies have shown that the E484K mutation, when introduced with N501Y mutation, can increase RBD-ACE2 binding. For instance, in the Beta and Gamma variants, the E484K-N501Y-D614G triple mutation has been found to enhance RBD-ACE2 binding by inducing local rearrangements involving rotamer placements between Q493 of RBD and H34 of ACE2 [75].

Two other RBD mutations, L452R and T478K, were also outlined as leading mutations in April 2020 and November 2020 by our deLemus analysis (Fig. 3). These mutations were found in the Delta variant that emerged in April 2021. Their locations within the epitope of several important neutralizing antibodies enable them to enhance the immune escape capabilities of the virus [69,76–78]. For the L452R mutation, computational studies have shown that variants possessing the L452R-E484Q-N501Y triple mutation exhibit a secondary structure rearrangement that is associated with an increase in RBD-ACE2 binding affinity [79]. For the T478K mutation, structural analysis has revealed that it allows the formation of two new hydrogen bonds located between Y489 of RBD and Y83 of ACE2, and F490 of RBD and K31 of ACE2, respectively, resulting in tighter binding between the RBD and ACE2 receptor [80].

The Omicron variant that emerged in November 2021, known for its exceptionally high transmissibility, harbors a significant number of RBD mutations. Most of these mutations have been reported in previous variants, but several new sites were outlined as leading mutations by our deLemus analysis, which encompasses mutations at R408, N440, and G446. The R408S mutation has been shown to alter the antigenic property of the spike glycoprotein by disrupting the binding of F2 antibodies [81]. Unlike R408S, the N440K mutation can enhance spike-ACE2 binding affinity by increasing the electrostatic complementarity between the structurally flexible RBM recognition site and the ACE2 receptor [82]. As for the G446 residue, which is situated at a highly antigenic region of the spike structure (Fig. 1), mutations have been shown to influence neutralization by both mAbs and antibodies present in polyclonal serum [77,78,83].

Overall, the leading mutations outlined within the RBD have been confirmed among the reported variants, suggesting their functional impact on viral fitness. Increased binding to the ACE2 receptor and immune evasion are factors that can enhance viral infectivity and overall viral fitness [84] [75]. These factors involve multiple amino acid residues located on the surface region of the spike protein. The spatial distribution analysis of the leading mutations highlights their prevalence on the surface region of the spike protein (Fig. 4), which can enhance viral infectivity through host interactions, such as improved ACE2 binding affinity and antibody escape. This observation is further supported by previously reported deep mutational scanning data, which assesses ACE2 binding affinity and antibody escape scores [70].

3.3. C-terminal domains (CTDs) and the S2 subunit

The post-RBD region of the SARS-CoV-2 spike glycoprotein consists of CTD1, CTD2, and the S2 subunit. While they do not directly engage the ACE2 receptor, they confer significant functions in spike allostery and membrane fusion [85,86]. The close spatial proximities between the CTDs and the NTD-RBD linker motif enable them to modulate RBD motion, and mutations in these regions could alter its open-close dynamics [87–89]. The S2 mediates virus-host membrane fusion

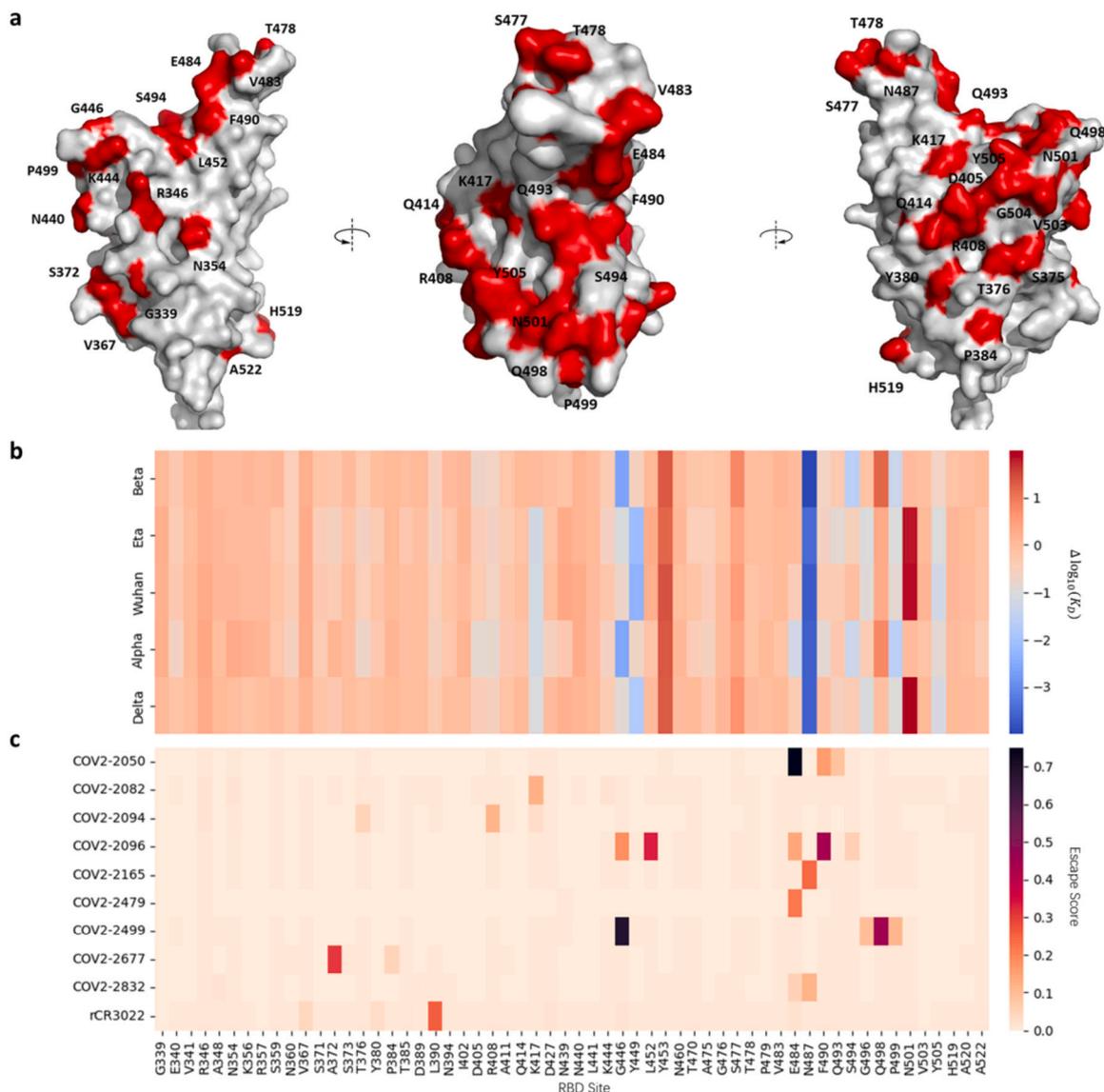


Fig. 4. Data analysis on deep mutational scanning from published work [70]. (a) 3D structure illustration of leading mutations outlined in RBD (red). (b) RBD-ACE2 binding affinity distribution with different reference virus variants. (c) Escape score of leading mutation sites.

by undergoing a cascade of conformational changes [23,86,90], and its mutations may impact spike fusogenicity both in terms of host-virus membrane fusion for cell surface or endosomal entry and cell-cell membrane fusion for syncytia formation [23,31]. Additionally, the post-RBD region houses the S1/S2 and S2' cleavage sites, which are proteolytically processed to facilitate membrane fusion [23,26,90–92]. The former site is generated by a polybasic insertion between CTD2 and S2_(681PRRAR↓S₆₈₆) unique to SARS-CoV-2, and has been found to enhance its infectivity [26,93,94]. Although their evolutionary importance is often overshadowed by those of the upstream NTD and RBD, the CTDs and S2 remain pivotal to the proper functioning of the spike.

The pre-Alpha stage of the pandemic marked the phase in which several mutations hallmarked to the CTDs of later variants emerged. As shown in Fig. 5, many of our CTD leading mutations outlined between January and December 2020 were reported in later VOCs and VOIs; these include: A570D, D614G, H655Y, Q677H, N679K, and P681H. One important leading mutation we outlined in the first month is D614G. It has been shown to promote RBD opening by disrupting an interprotomer hydrogen bond involving S2 residues [95], and to increase S1/S2 cleavage, cell entry, replicative fitness, and transmissibility [95–97]. These enhancements in fitness led to its fixation in the global SARS-CoV-

2 population. Most of the CTD mutations outlined above follow a similar trend in either finetuning RBD opening or cleavage efficiency. For example, A570D of Alpha has been revealed to form new interprotomer hydrogen bonds and salt-bridge interactions with S2 residues to modulate RBD motion [85,89]. Meanwhile, at the highly polymorphic P681 site, both P681H and P681R has been reported to increase S1/S2 cleavage efficiency [98–100]. However, only P681R of Delta would have a pronounced effect in improving spike fusogenicity [98–100]. Interestingly, the co-occurrence of N679K and P681H in Omicron variants has been found to introduce a novel cathepsin G cleavage site proximal to the S1/S2 furin cleavage site [101], which may explain their switch from cell surface entry pathways preferred by pre-Omicron strains to endosomal entry pathways [102–106].

Our deLemus method is also effective in outlining S2 mutations. As shown in Fig. 6, between January and December 2020, several outlined leading mutations of the S2 subunit were later confirmed in subsequent variants; these are: A701V, T716I, D796Y, S982A, D1118H, and V1176F. Unlike CTD mutations, the effects of the S2 mutations are more diverse [107]. Both T716I and S982A of Alpha have been structurally determined to confer local destabilizing effect [85]. In particular, the latter mutation has been shown to abrogate an interprotomer hydrogen

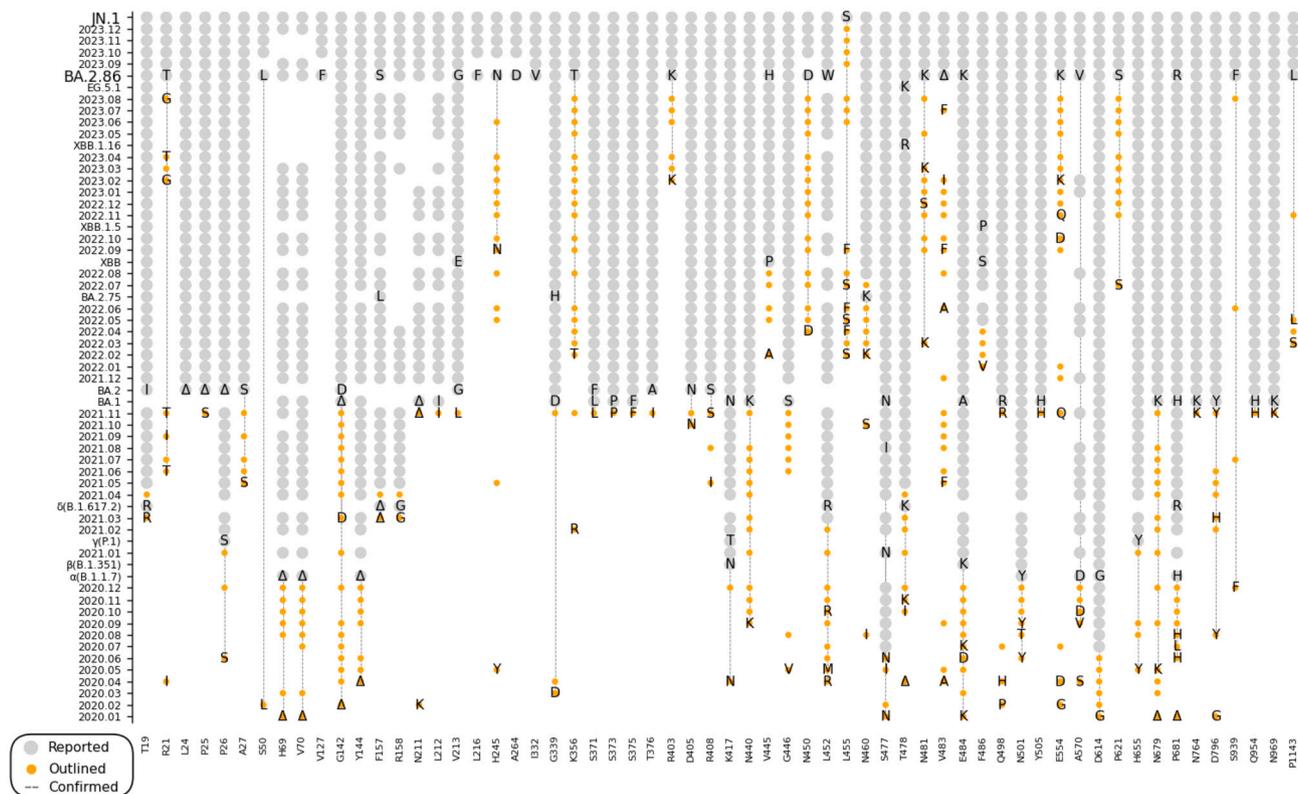


Fig. 7. Evolutionary trajectory of the BA.2.86 and JN.1. The annotations, labels, and color schemes in this figure are identical to those employed in Fig. 2.

ing beyond the 24-month timeframe discussed in the preceding sections. Any leading mutations that were subsequently confirmed in circulating variants have been annotated and updated on the deLemus webpage. It is worth noting that certain mutations consistently appeared in our analysis. As shown in Fig. 7, the persistent post-RBD mutations are: E554K, A570V, P621S, and S939F. These mutations would later be reported in the BA.2.86 lineage of Omicron which emerged in August 2023. This variant harbors over 30 new spike mutations when compared to the parental BA.2 strain [117,118]. Recent characterization has revealed that BA.2.86 shows improved ACE2-binding affinity [119], as well as enhanced immune escape capabilities due to its distinct antigenic profile [118,120–122]. We also identified persistent mutations in the RBD region, namely K356T, R403K, and N450D, which are confirmed in BA.2.86 (Fig. 7). Lastly, from the beginning of 2022, we observed a highly polymorphic site, L455, corresponding to the L455F/S mutations (Fig. 7). In particular, L455S is a novel mutation carried by JN.1, a descendant lineage of BA.2.86, which emerged in December 2023. Remarkably, even though the two variants only differ by this single mutation, JN.1 has been demonstrated to exhibit even better immune evasion than BA.2.86 [123,124]. This has likely enabled it to become the most prevalent strain globally since February 2024 [125].

In addition to outlining the leading mutations, we revealed some general evolutionary features of the SARS-CoV-2 spike glycoprotein in terms of Ξ and \bar{s} . Mutations are the source of genetic variation, and how mutation rates fluctuate over the course of evolution is of particular interest [126]. Based on their Ξ values, three characteristic evolutionary phases can be distinguished (Fig. S2). The first phase lasted from December 2019 to October 2020, when Ξ maintained a steady state at relatively low values. In November 2020, a month before the emergence of Alpha, the second phase began with an increase of Ξ , to a maximum in March 2021, after which Delta appeared. This is followed by the third phase, marking the gradual decrease of Ξ back to a steady state by June 2021. As most mutations are assumed to be neutral or slightly deleterious [14–16], selection often acts against high mutation rates [127], which is illustrated by the initially low Ξ . This may also im-

ply a dynamic equilibrium within the viral quasispecies [37,128,129], where each pre-Alpha variant has similar fitness. The subsequent rise in Ξ is thought to be a consequence of environmental changes [37], which would increase the chances for beneficial mutations to occur [127]. However, more deleterious mutations would likewise be introduced [127,130]. This may explain the resulting drop in Ξ . On the other hand, genetic diversity measured by \bar{s} exhibited a clear increasing trend from December 2019 (Fig. S2). This suggests the ongoing adaptive evolution of the spike [131], although further research would be required to elucidate the exact mechanisms that govern the complex evolutionary trajectory of SARS-CoV-2.

4. Conclusion

This work introduces the novel deLemus method for analyzing the evolutionary dynamics of the SARS-CoV-2 spike glycoprotein. Our proposed L -index is effectual in quantifying the mutation strength of each amino acid site, such that leading mutations can be outlined (Table S2). Comprehensive characterization of these leading mutations reveals how the spike’s complex mutation pattern is shaped by the distinct evolutionary trajectory of each of its functional domains: the antigenic evolution of the NTD, the dynamic balance between ACE2-binding affinity and immune evasion by RBD mutations, and the cell entry mechanism shift modulated by post-RBD mutations. With its effectiveness in systematically monitoring the ongoing SARS-CoV-2 evolution, it may be feasible to extend our deLemus method to analyze other viral evolution processes. The ability to identify a subset of evolutionarily significant sites in circulating viruses would accelerate the development of treatments and disease control measures in the case of future pandemics.

CRedit authorship contribution statement

Muhammad Hasan: Data curation, Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing – original

draft, Writing – review & editing. **Zhouyi He**: Formal analysis, Investigation, Methodology, Validation, Writing – original draft, Writing – review & editing. **Mengqi Jia**: Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. **Alvin C.F. Leung**: Formal analysis, Investigation, Validation, Writing – original draft, Writing – review & editing. **Kathiresan Natarajan**: Formal analysis, Investigation, Validation, Writing – original draft, Writing – review & editing. **Wentao Xu**: Investigation, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. **Shanqi Yap**: Data curation, Formal analysis, Investigation, Validation, Writing – original draft, Writing – review & editing. **Feng Zhou**: Investigation, Validation, Writing – original draft, Writing – review & editing. **Shihong Chen**: Data curation, Investigation, Software, Validation, Writing – original draft, Writing – review & editing. **Hailei Su**: Investigation, Validation, Writing – original draft, Writing – review & editing. **Kaicheng Zhu**: Formal analysis, Investigation, Methodology, Validation, Writing – original draft, Writing – review & editing. **Haibin Su**: Conceptualization, Funding acquisition, Investigation, Methodology, Project administration, Supervision, Validation, Writing – original draft, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

All input data are freely available from GISAID. The code employed in this study is publicly accessible via the GitHub repository at <https://github.com/hbsulab/L-index.git>. The results are available on the website and regularly updated on a monthly basis at <https://hbsulab.github.io/deLemus/>.

Acknowledgements

We acknowledge the data contributors and the GISAID Initiative for sharing the genetic sequence. We are grateful to Professor Yong Huang for many thought-provoking discussions and encouragement in the course of this work, which is supported in part by the Hong Kong Branch of South Marine Science and Engineering Guangdong Laboratory (SMSEGL20SC01) (for H.S.), ASPIRE League Partnership Seed Fund Program (ASPIRE 2019#1), and Society of Interdisciplinary Research (SOIRÉE). This work has been funded by the Research Grants Council of Hong Kong (16305623) and HKUST grant (R9418). This research made use of the computing resources of the X-GPU cluster supported by the Hong Kong Research Grants Council Collaborative Research Fund (C6021-19EF). A.C.F.L. and S.Q.Y. are grateful for the support from Undergraduate Research Opportunities Program (UROP) at HKUST. We would also like to acknowledge the enlightening advice from Jie Gao in the early stage of this work; the technical support from Long T. Ta and Siqi Lei for improving figures; and Ruichen Li, Leon R. Salim, and Salina Lam in data collection and preparation.

Appendix A. Supplementary material

Supplementary material related to this article can be found online at <https://doi.org/10.1016/j.csbj.2024.05.037>.

References

- [1] Zhou P, Yang X-L, Wang X-G, et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* 2020;579:270–3.
- [2] WHO Coronavirus (COVID-19) Dashboard. <https://covid19.who.int/>.

- [3] The World Health Organization. WHO Director-General's opening remarks at the media briefing – 5 May 2023. <https://www.who.int/news-room/speeches/item/who-director-general-s-opening-remarks-at-the-media-briefing-5-may-2023>, 2023.
- [4] Minskaia E, Hertzog T, Gorbalenya AE, et al. Discovery of an RNA virus 3 → 5 exoribonuclease that is critically involved in coronavirus RNA synthesis. *Proc Natl Acad Sci USA* 2006;103:5108.
- [5] Denison MR, Graham RL, Donaldson EF, et al. Coronaviruses: an RNA proofreading machine regulates replication fidelity and diversity. *RNA Biol* 2011;8:270–9.
- [6] Liu C, Shi W, Becker ST, et al. Structural basis of mismatch recognition by a SARS-CoV-2 proofreading enzyme. *Science* 2021;373:1142–6.
- [7] Moeller NH, Shi K, Demir Özlem, et al. Structure and dynamics of SARS-CoV-2 proofreading exoribonuclease ExoN. *Proc Natl Acad Sci USA* 2022;119:e2106379119.
- [8] McCarthy KR, Rennick LJ, Nambulli S, et al. Recurrent deletions in the SARS-CoV-2 spike glycoprotein drive antibody escape. *Science* 2021;371:1139–42.
- [9] Li X, Giorgi EE, Marichannegowda MH, et al. Emergence of SARS-CoV-2 through recombination and strong purifying selection. *Sci Adv* 2020;6:eabb9153.
- [10] Carabelli AM, Peacock TP, Thorne LG, et al. SARS-CoV-2 variant biology: immune escape, transmission and fitness. *Nat Rev Microbiol* 2023.
- [11] Markov PV, Ghafari M, Beer M, et al. The evolution of SARS-CoV-2. *Nat Rev Microbiol* 2023;21:361–79.
- [12] Rambaut A, Holmes EC, O'Toole Á, et al. A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. *Nat Microbiol* 2020;5:1403–7.
- [13] Tracking SARS-CoV-2 variants. <https://www.who.int/en/activities/tracking-SARS-CoV-2-variants/>.
- [14] Kimura M. Evolutionary rate at the molecular level. *Nature* 1968;217:624–6.
- [15] Ohta T. Slightly deleterious mutant substitutions in evolution. *Nature* 1973;246:96–8.
- [16] Nei M. Selectionism and neutralism in molecular evolution. *Mol Biol Evol* 2005;22:2318–42.
- [17] Orr HA. The population genetics of beneficial mutations. *Philos Trans R Soc B* 2010;365:1195–201.
- [18] Harvey WT, Carabelli AM, Jackson B, et al. SARS-CoV-2 variants, spike mutations and immune escape. *Nat Rev Microbiol* 2021;19:409–24.
- [19] McCormick KD, Jacobs JL, Mellors JW. The emerging plasticity of SARS-CoV-2. *Science* 2021;371:1306–8.
- [20] Tiwari M, Mishra D. Investigating the genomic landscape of novel coronavirus (2019-nCoV) to identify non-synonymous mutations for use in diagnosis and drug design. *J Clin Virol* 2020;128:104441.
- [21] Mishra D, Suri GS, Kaur G, et al. Comparative insight into the genomic landscape of SARS-CoV-2 and identification of mutations associated with the origin of infection and diversity. *J Med Virol* 2021;93:2406–19.
- [22] Saldivar-Espinoza B, Garcia-Segura P, Novau-Ferré N, et al. The mutational landscape of SARS-CoV-2. *Int J Mol Sci* 2023;24.
- [23] Jackson CB, Farzan M, Chen B, et al. Mechanisms of SARS-CoV-2 entry into cells. *Nat Rev Mol Cell Biol* 2022;23:3–20.
- [24] Renhong Y, Yuanyuan Z, Yaning L, et al. Structural basis for the recognition of SARS-CoV-2 by full-length human ACE2. *Science* 2020;367:1444–8.
- [25] Benton DJ, Wrobel AG, Xu P, et al. Receptor binding and priming of the spike protein of SARS-CoV-2 for membrane fusion. *Nature* 2020;588:327–30.
- [26] Hoffmann M, Kleine-Weber H, Pöhlmann S. A multibasic cleavage site in the spike protein of SARS-CoV-2 is essential for infection of human lung cells. *Mol Cell* 2020;78:779–784.e.
- [27] Walls AC, Park Y-J, Tortorici MA, et al. Structure, function, and antigenicity of the SARS-CoV-2 spike glycoprotein. *Cell* 2020;181:281–292.e.
- [28] Hansen J, Baum A, Pascal KE, et al. Studies in humanized mice and convalescent humans yield a SARS-CoV-2 antibody cocktail. *Science* 2020;369:1010–4.
- [29] Cohen KW, Linderman SL, Moodie Z, et al. Longitudinal analysis shows durable and broad immune memory after SARS-CoV-2 infection with persisting antibody responses and memory B and T cells. *Cell Rep Med* 2021;2:100354.
- [30] Moss P. The T cell immune response against SARS-CoV-2. *Nat Immunol* 2022;23:186–93.
- [31] Huang Y, Yang C, Xu X-f, et al. Structural and functional properties of SARS-CoV-2 spike protein: potential antiviral drug development for COVID-19. *Acta Pharmacol Sin* 2020;41:1141–9.
- [32] Dai L, Gao GF. Viral targets for vaccines against COVID-19. *Nat Rev Immunol* 2021;21:73–82.
- [33] Su H, Zhou F, Huang Z, et al. Molecular insights into small-molecule drug discovery for SARS-CoV-2. *Angew Chem Int Ed* 2021;60:9789–802.
- [34] Chen RE, Zhang X, Case JB, et al. Resistance of SARS-CoV-2 variants to neutralization by monoclonal and serum-derived polyclonal antibodies. *Nat Med* 2021;27:717–26.
- [35] Shu Y, McCauley J. GISAID: global initiative on sharing all influenza data – from vision to reality. *Euro Surveill* 2017;22:30494.
- [36] Worth CL, Gong S, Blundell TL. Structural and functional constraints in the evolution of protein families. *Nat Rev Mol Cell Biol* 2009;10:709–20.
- [37] Domingo E, Sheldon J, Perales C. Viral quasispecies evolution. *Microbiol Mol Biol Rev* 2012;76:159–216.

- [38] Jaroszewski L, Iyer M, Alisoltani A, et al. The interplay of SARS-CoV-2 evolution and constraints imposed by the structure and functionality of its proteins. *PLoS Comput Biol* 2021;17:1–22.
- [39] Tenthorey JL, Emerman M, Malik HS. Evolutionary landscapes of host-virus arms races. *Annu Rev Immunol* 2022;40:271–94.
- [40] Lauring AS, Andino R. Quasispecies theory and the behavior of RNA viruses. *PLoS Pathog* 2010;6:1–8.
- [41] Bazykin GA. Changing preferences: deformation of single position amino acid fitness landscapes and evolution of proteins. *Biol Lett* 2015;11:20150315.
- [42] Delport W, Scheffler K, Seoghe C. Frequent toggling between alternative amino acids is driven by selection in HIV-1. *PLoS Pathog* 2008;4:1–13.
- [43] Borucki MK, Allen JE, Chen-Harris H, et al. The role of viral population diversity in adaptation of bovine coronavirus to new host environments. *PLoS ONE* 2013;8:1–11.
- [44] Sheftel H, Szekely P, Mayo A, et al. Evolutionary trade-offs and the structure of polymorphisms. *Philos Trans R Soc B* 2018;373:20170105.
- [45] Gao R, Zu W, Liu Y, et al. Quasispecies of SARS-CoV-2 revealed by single nucleotide polymorphisms (SNPs) analysis. *Virulence* 2021;12:1209–26.
- [46] Wu F, Zhao S, Yu B, et al. A new coronavirus associated with human respiratory disease in China. *Nature* 2020;579:265–9.
- [47] Goujon M, McWilliam H, Li W, et al. A new bioinformatics analysis tools framework at EMBL–EBI. *Nucleic Acids Res* 2010;38:W695–9.
- [48] Golub GH, Van Loan CF. *Matrix computations*. 4th ed. Maryland: The Johns Hopkins University Press; 2013. p. 780.
- [49] Jumper J, Evans R, Pritzel A, et al. Highly accurate protein structure prediction with AlphaFold. *Nature* 2021;596:583–9.
- [50] Wrapp D, Wang N, Corbett KS, et al. Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. *Science* 2020;367:1260–3.
- [51] Suryadevara N, Shrihari S, Gilchuk P, et al. Neutralizing and protective human monoclonal antibodies recognizing the N-terminal domain of the SARS-CoV-2 spike protein. *Cell* 2021;184:2316–2331.e5.
- [52] Wang Z, Muecksch F, Cho A, et al. Analysis of memory B cells identifies conserved neutralizing epitopes on the N-terminal domain of variant SARS-CoV-2 spike proteins. *Immunity* 2022;55:998–1012.e8.
- [53] Qing E, Kicmal T, Kumar B, et al. Dynamics of SARS-CoV-2 spike proteins in cell entry: control elements in the amino-terminal domains. *mBio* 2021;12:e01590-21.
- [54] Ray D, Le L, Andricioaei I. Distant residues modulate conformational opening in SARS-CoV-2 spike protein. *Proc Natl Acad Sci USA* 2021;118:e2100943118.
- [55] Alisoltani A, Jaroszewski L, Iyer M, et al. Increased frequency of indels in hyper-variable regions of SARS-CoV-2 proteins—a possible signature of adaptive selection. *Front Genet* 2022;13.
- [56] Weng S, Zhou H, Ji C, et al. Conserved pattern and potential role of recurrent deletions in SARS-CoV-2 evolution. *Microbiol Spectr* 2022;10:e02191-21.
- [57] Meng B, Kemp SA, Papa G, et al. Recurrent emergence of SARS-CoV-2 spike deletion H69/V70 and its role in the Alpha variant B.1.1.7. *Cell Rep* 2021;35:109292.
- [58] Chi X, Yan R, Zhang J, et al. A neutralizing human antibody binds to the N-terminal domain of the Spike protein of SARS-CoV-2. *Science* 2020;369:650–5.
- [59] McCallum M, De Marco A, Lempp FA, et al. N-terminal domain antigenic mapping reveals a site of vulnerability for SARS-CoV-2. *Cell* 2021;184:2332–2347.e6.
- [60] Wibmer CK, Ayres F, Hermanus T, et al. SARS-CoV-2 501Y. V2 escapes neutralization by South African COVID-19 donor plasma. *Nat Med* 2021;27:622–5.
- [61] Dejnirattisai W, Zhou D, Supasa P, et al. Antibody evasion by the P.1 strain of SARS-CoV-2. *Cell* 2021;184:2939–2954.e9.
- [62] Chaudhari AM, Joshi M, Kumar D, et al. Evaluation of immune evasion in SARS-CoV-2 Delta and Omicron variants. *Comput Struct Biotechnol J* 2022;20:4501–16.
- [63] Liu L, Iketani S, Guo Y, et al. Striking antibody evasion manifested by the Omicron variant of SARS-CoV-2. *Nature* 2022;602:676–81.
- [64] Tai W, He L, Zhang X, et al. Characterization of the receptor-binding domain (RBD) of 2019 novel coronavirus: implication for development of RBD protein as a viral attachment inhibitor and vaccine. *Cell Mol Immunol* 2020;17:613–20.
- [65] Lan J, Ge J, Yu J, et al. Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. *Nature* 2020;581:215–20.
- [66] Premkumar L, Segovia-Chumbez B, Jadi R, et al. The receptor-binding domain of the viral spike protein is an immunodominant and highly specific target of antibodies in SARS-CoV-2 patients. *Sci Immunol* 2020;5:eabc8413.
- [67] Piccoli L, Park Y-J, Tortorici MA, et al. Mapping neutralizing and immunodominant sites on the SARS-CoV-2 spike receptor-binding domain by structure-guided high-resolution serology. *Cell* 2020;183:1024–1042.e21.
- [68] Tao K, Tzou PL, Nounin J, et al. The biological and clinical significance of emerging SARS-CoV-2 variants. *Nat Rev Genet* 2021;22:757–73.
- [69] Wang Z, Schmidt F, Weisblum Y, et al. mRNA vaccine-elicited antibodies to SARS-CoV-2 and circulating variants. *Nature* 2021;592:616–22.
- [70] Starr TN, Greaney AJ, Hilton SK, et al. Deep mutational scanning of SARS-CoV-2 receptor binding domain reveals constraints on folding and ACE2 binding. *Cell* 2020;182:1295–1310.e20.
- [71] Weisblum Y, Schmidt F, Zhang F, et al. Escape from neutralizing antibodies by SARS-CoV-2 spike protein variants. *eLife* 2020;9.
- [72] Greaney AJ, Loes AN, Crawford KHD, et al. Comprehensive mapping of mutations in the SARS-CoV-2 receptor-binding domain that affect recognition by polyclonal human plasma antibodies. *Cell Host Microbe* 2021;29:463–476.e6.
- [73] Baum A, Fulton BO, Wloga E, et al. Antibody cocktail to SARS-CoV-2 spike protein prevents rapid mutational escape seen with individual antibodies. *Science* 2020;369:1014–8.
- [74] Saville JW, Mannar D, Zhu X, et al. Structural and biochemical rationale for enhanced spike protein fitness in delta and kappa SARS-CoV-2 variants. *Nat Commun* 2022;13:742.
- [75] Mannar D, Saville JW, Zhu X, et al. Structural analysis of receptor binding domain mutations in SARS-CoV-2 variants of concern that modulate ACE2 and antibody binding. *Cell Rep* 2021;37:110156.
- [76] Planas D, Veyer D, Baidaliuk A, et al. Reduced sensitivity of SARS-CoV-2 variant Delta to antibody neutralization. *Nature* 2021;596(7871):276–80.
- [77] Li Q, Wu J, Nie J, et al. The impact of mutations in SARS-CoV-2 spike on viral infectivity and antigenicity. *Cell* 2020;182:1284–1294.e9.
- [78] Liu Z, VanBlargan LA, Bloyet L-M, et al. Identification of SARS-CoV-2 spike mutations that attenuate monoclonal and serum antibody neutralization. *Cell Host Microbe* 2021;29:477–488.e4.
- [79] Aggarwal A, Naskar S, Maroli N, et al. Mechanistic insights into the effects of key mutations on SARS-CoV-2 RBD–ACE2 binding. *Phys Chem Chem Phys* 2021;23(46):26451–8.
- [80] Wang Y, Liu C, Zhang C, et al. Structural basis for SARS-CoV-2 Delta variant recognition of ACE2 receptor and broadly neutralizing antibodies. *Nat Commun* 2022;13:871.
- [81] Cao Y, Yisimayi A, Jian F, et al. BA.2.12.1, BA.4 and BA.5 escape antibodies elicited by Omicron infection. *Nature* 2022;608:593–602.
- [82] Gan HH, Twaddle A, Marchand B, et al. Structural modeling of the SARS-CoV-2 spike/human ACE2 complex interface can identify high-affinity variants associated with increased transmissibility. *J Mol Biol* 2021;433:167051.
- [83] Starr TN, Greaney AJ, Addetia A, et al. Prospective mapping of viral mutations that escape antibodies used to treat COVID-19. *Science* 2021;371:850–4.
- [84] Yi C, Sun X, Ling Z, et al. Jigsaw puzzle of SARS-CoV-2 RBD evolution and immune escape. *Cell Mol Immunol* 2022;19:848–51.
- [85] Gobeil SM-C, Janowska K, McDowell S, et al. Effect of natural mutations of SARS-CoV-2 on spike structure, conformation, and antigenicity. *Science* 2021;373:eabi6226.
- [86] Fan X, Cao D, Kong L, et al. Cryo-EM analysis of the post-fusion structure of the SARS-CoV spike glycoprotein. *Nat Commun* 2020;11:3618.
- [87] Henderson R, Edwards RJ, Mansouri K, et al. Controlling the SARS-CoV-2 spike glycoprotein conformation. *Nat Struct Mol Biol* 2020;27:925–33.
- [88] Gobeil SM-C, Janowska K, McDowell S, et al. D614G mutation alters SARS-CoV-2 spike conformation and enhances protease cleavage at the S1/S2 junction. *Cell Rep* 2021;34:108630.
- [89] Yang T-J, Yu P-Y, Chang Y-C, et al. Effect of SARS-CoV-2 B.1.1.7 mutations on spike protein structure and function. *Nat Struct Mol Biol* 2021;28:731–9.
- [90] Yu S, Zheng X, Zhou B, et al. SARS-CoV-2 spike engagement of ACE2 primes S2 site cleavage and fusion initiation. *Proc Natl Acad Sci USA* 2022;119:e211199119.
- [91] Belouzard S, Chu VC, Whittaker GR. Activation of the SARS coronavirus spike protein via sequential proteolytic cleavage at two distinct sites. *Proc Natl Acad Sci USA* 2009;106:5871–6.
- [92] Rajah MM, Bernier A, Buchrieser J, et al. The mechanism and consequences of SARS-CoV-2 spike-mediated fusion and syncytia formation. *J Mol Biol* 2022;434:167280.
- [93] Jaimes JA, Millet JK, Whittaker GR. Proteolytic cleavage of the SARS-CoV-2 spike protein and the role of the novel S1/S2 site. *iScience* 2020;23:101212.
- [94] Papa G, Mallery DL, Albecka A, et al. Furin cleavage of SARS-CoV-2 Spike promotes but is not essential for infection and cell-cell fusion. *PLoS Pathog* 2021;17:1–20.
- [95] Gobeil SM-C, Janowska K, McDowell S, et al. D614G mutation alters SARS-CoV-2 Spike conformation and enhances protease cleavage at the S1/S2 junction. *Cell Rep* 2021;34.
- [96] Ozono S, Zhang Y, Ode H, et al. SARS-CoV-2 D614G spike mutation increases entry efficiency with enhanced ACE2-binding affinity. *Nat Commun* 2021;12:848.
- [97] Zhou B, Thao TTN, Hoffmann D, et al. SARS-CoV-2 spike D614G change enhances replication and transmission. *Nature* 2021;592:122–7.
- [98] Lubinski B, Fernandes MH, Frazier L, et al. Functional evaluation of the P681H mutation on the proteolytic activation of the SARS-CoV-2 variant B.1.1.7 (Alpha) spike. *iScience* 2022;25:103589.
- [99] Saito A, Irie T, Suzuki R, et al. Enhanced fusogenicity and pathogenicity of SARS-CoV-2 Delta P681R mutation. *Nature* 2022;602:300–6.
- [100] Liu Y, Liu J, Johnson BA, et al. Delta spike P681R mutation enhances SARS-CoV-2 fitness over Alpha variant. *Cell Rep* 2022;39.
- [101] Mustafa Z, Kalbacher H, Burster T. Occurrence of a novel cleavage site for cathepsin G adjacent to the polybasic sequence within the proteolytically sensitive activation loop of the SARS-CoV-2 Omicron variant: the amino acid substitution N679K and P681H of the spike protein. *PLoS ONE* 2022;17:1–11.
- [102] Meng B, Abdullahi A, Ferreira IATM, et al. Altered TMPRSS2 usage by SARS-CoV-2 Omicron impacts infectivity and fusogenicity. *Nature* 2022;603:706–14.
- [103] Du X, Tang H, Gao L, et al. Omicron adopts a different strategy from Delta and other variants to adapt to host. *Signal Transduct Targeted Ther* 2022;7:45.
- [104] Hui KPY, Ho JCW, Cheung M-c, et al. SARS-CoV-2 Omicron variant replication in human bronchus and lung ex vivo. *Nature* 2022;603:715–20.
- [105] Willett BJ, Grove J, MacLean OA, et al. SARS-CoV-2 Omicron is an immune escape variant with an altered cell entry pathway. *Nat Microbiol* 2022;7:1161–79.

- [106] Shuai H, Chan JF-W, Hu B, et al. The viral fitness and intrinsic pathogenicity of dominant SARS-CoV-2 Omicron sublineages BA.1, BA.2, and BA.5. *eBioMedicine* 2023;95.
- [107] Magazine N, Zhang T, Wu Y, et al. Mutations and evolution of the SARS-CoV-2 Spike protein. *Viruses* 2022;14.
- [108] Xia S, Wen Z, Wang L, et al. Structure-based evidence for the enhanced transmissibility of the dominant SARS-CoV-2 B.1.1.7 variant (Alpha). *Cell Discov* 2021;7:109.
- [109] Kumar S, Delipan R, Chakraborty D, et al. Mutations in S2 subunit of SARS-CoV-2 Omicron spike strongly influence its conformation, fusogenicity, and neutralization sensitivity. *J Virol* 2023;97:e00922-23.
- [110] Elko EA, Mead HL, Nelson GA, et al. Recurrent SARS-CoV-2 mutations at Spike D796 evade antibodies from pre-Omicron convalescent and vaccinated subjects. *Microbiol Spectr* 2024;12:e03291-23.
- [111] Yang K, Wang C, White KI, et al. Structural conservation among variants of the SARS-CoV-2 spike postfusion bundle. *Proc Natl Acad Sci USA* 2022;119:e2119467119.
- [112] Zhang J, Cai Y, Lavine CL, et al. Structural and functional impact by SARS-CoV-2 Omicron spike mutations. *Cell Rep* 2022;39:110729.
- [113] Sun C, Wang H, Yang J, et al. Mutation N856K in spike reduces fusogenicity and infectivity of Omicron BA.1. *Signal Transduct Targeted Ther* 2023;8:75.
- [114] Park SB, Khan M, Chiliveri SC, et al. SARS-CoV-2 omicron variants harbor spike protein mutations responsible for their attenuated fusogenic phenotype. *Commun Biol* 2023;6:556.
- [115] Escalera A, Laporte M, Turner S, et al. The impact of S2 mutations on Omicron SARS-CoV-2 cell surface expression and fusogenicity. *Emerg Microbes Infect* 2024;13:2297553.
- [116] Furusawa Y, Kiso M, Iida S, et al. In SARS-CoV-2 delta variants, Spike-P681R and D950N promote membrane fusion, Spike-P681R enhances spike cleavage, but neither substitution affects pathogenicity in hamsters. *eBioMedicine* 2023;91.
- [117] Looi M-K. Covid-19: scientists sound alarm over new BA.2.86 “Pirola” variant. *BMJ* 2023;382.
- [118] Uriu K, Ito J, Kosugi Y, et al. Transmissibility, infectivity, and immune evasion of the SARS-CoV-2 BA.2.86 variant. *Lancet Infect Dis* 2023;23:e460–1.
- [119] Wang Q, Guo Y, Liu L, et al. Antigenicity and receptor affinity of SARS-CoV-2 BA.2.86 spike. *Nature* 2023;624:639–44.
- [120] Yang S, Yu Y, Jian F, et al. Antigenicity and infectivity characterisation of SARS-CoV-2 BA.2.86. *Lancet Infect Dis* 2023;23:e457–9.
- [121] Lasrado N, Collier A-RY, Hachmann NP, et al. Neutralization escape by SARS-CoV-2 Omicron subvariant BA.2.86. *Vaccine* 2023;41:6904–9.
- [122] Khan K, Lustig G, Römer C, et al. Evolution and neutralization escape of the SARS-CoV-2 BA.2.86 subvariant. *Nat Commun* 2023;14:8078.
- [123] Yang S, Yu Y, Xu Y, et al. Fast evolution of SARS-CoV-2 BA.2.86 to JN.1 under heavy immune pressure. *Lancet Infect Dis* 2023;3099:10–2.
- [124] Yamasoba D, Uriu K, Plianachaisuk A, et al. Virological characteristics of the SARS-CoV-2 JN.1 variant. *Lancet Infect Dis* 2023;3099:3–4.
- [125] The World Health Organization. 9 February 2024 JN.1 Updated Risk Evaluation. https://www.who.int/docs/default-source/coronaviruse/09022024_jn_1_ure.pdf?sfvrsn=a153518c_3_2024.
- [126] Lynch M. Evolution of the mutation rate. *Trends Genet* 2010;26:345–52.
- [127] Peck KM, Luring AS. Complexities of viral mutation rates. *J Virol* 2018;92:e01031-17.
- [128] Domingo E, Sabo D, Taniguchi T, et al. Nucleotide sequence heterogeneity of an RNA phage population. *Cell* 1978;13:735–44.
- [129] Steinhauer DA, Holland JJ. Rapid evolution of RNA viruses. *Annu Rev Microbiol* 1987;41:409–31.
- [130] Couce A, Caudwell LV, Feinauer C, et al. Mutator genomes decay, despite sustained fitness gains, in a long-term experiment with bacteria. *Proc Natl Acad Sci USA* 2017;114:E9026–35.
- [131] Ellegren H, Galtier N. Determinants of genetic diversity. *Nat Rev Genet* 2016;17:422–33.