Database, 2025, baaf002
DOI: https://doi.org/10.1093/database/baaf002
Database tool

AnnCovDB: a manually curated annotation database for mutations in SARS-CoV-2 spike protein

Xiaomin Zhang[1], Zhongyi Lei[1,2], Jiarong Zhang[1,3], Tingting Yang[1,3], Xian Liu[1], Jiguo Xue[1,*], Ming Ni[1,*]

[1]Academy of Military Medical Sciences, No. 27 Taiping Road, Haidian District, Beijing 100850, PR China
[2]College of Life Science and Technology, Beijing University of Chemical Technology, No.15 North Third Ring Road East, Chaoyang District, Beijing 100029, PR China
[3]School of Forensic Medicine, Shanxi Medical University, No.98, University Street, Wujinshan Town, Yuci District, Jinzhong, Shanxi Province 030600, PR China

[*]Corresponding authors. Ming Ni, Academy of Military Medical Sciences, Beijing 100850, PR China. E-mail: niming@bmi.ac.cn; Jiguo Xue, Academy of Military Medical Sciences, Beijing 100850, PR China. E-mail: xuejgcn@163.com.

Citation details: Zhang, X., Lei, Z., Zhang, J. *et al.* AnnCovDB: a manually curated annotation database for mutations in SARS-CoV-2 spike protein. *Database* (2025) Vol. 2025: article ID baaf002; DOI: https://doi.org/10.1093/database/baaf002

Abstract

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) has been circulating and adapting within the human population for >4 years. A large number of mutations have occurred in the viral genome, resulting in significant variants known as variants of concern (VOCs) and variants of interest (VOIs). The spike (S) protein harbors many of the characteristic mutations of VOCs and VOIs, and significant efforts have been made to explore functional effects of the mutations in the S protein, which can cause or contribute to viral infection, transmission, immune evasion, pathogenicity, and illness severity. However, the knowledge and understanding are dispersed throughout various publications, and there is a lack of a well-structured database for functional annotation that is based on manual curation. AnnCovDB is a database that provides manually curated functional annotations for mutations in the S protein of SARS-CoV-2. Mutations in the S protein carried by at least 8000 variants in the GISAID were chosen, and the mutations were then utilized as query keywords to search in the PubMed database. The searched publications revealed that 2093 annotation entities for 205 single mutations and 93 multiple mutations were manually curated. These entities were organized into multilevel hierarchical categories for user convenience. For example, one annotation entity of N501Y mutation was 'Infectious cycle ➔Attachment ➔ACE2 binding affinity ➔Increase'. AnnCovDB can be used to query specific mutations and browse through function annotation entities.
Database URL: https://AnnCovDB.app.bio-it.tech/

Introduction

By June 2024, the number of coronavirus disease 2019 (COVID-19) cases and deaths reported to the World Health Organization was over 775 million and caused 7 million [1]. The COVID-19 viral pathogen, severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), has highly diverse genomes and has evolved into about 2000 lineages [2]. Specifically, the variants of the Omicron lineage have dramatically more mutations in the spike (S) protein compared to those of previous variants of concern [3]. The circulation and adaption of the Omicron variants in the human population is continuing, resulting in thousands of deaths worldwide each month [1].

The S protein of SARS-CoV-2 binds to the angiotensin-converting enzyme 2 (ACE2) receptor on the surface of human cells. This binding initiates the entry of viral particles into the cell through TMPRSS2 or other endosomal proteases, which are associated with the non-Omicron and Omicron variants, respectively [4–7]. The S protein is essential for SARS-CoV-2, and mutations in the protein that improve SARS-CoV-2 fitness occur continuously [8]. For example, the D614G mutation in the S protein emerged in February 2020, preventing premature S1 shedding and thereby increasing infectivity [9–13]. N501Y significantly increases the affinity of the S protein for ACE2, which is amplified by epistasis interaction with the Q498R mutation [14–17]. The P681R mutation in the furin cleavage site of the Delta variant increased fusogenicity and pathogenicity [18–20]. Many mutations, including E484K/Q and L452R, have been shown to confer immune evasion capabilities, particularly for the Omicron variants [21, 22].

While many studies have been focused on the functions of S protein mutations, there is currently a paucity of databases that encompass relevant publications and provide a thorough integration of functional annotations. The COV2Var database, developed by Feng *et al.*, provides a comprehensive computational assessment of the effects of SARS-CoV-2 mutations and their relationships with various factors

Received 25 July 2024; Revised 5 January 2025; Accepted 7 February 2025
© The Author(s) 2025. Published by Oxford University Press.
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

[23]; however, publications queried by specific mutations are shown automatically, without manual curation. The CoVariants (https://covariants.org) database and the Stanford Coronavirus Resistance Database (CoV-RDB) both provide manually curated annotations for mutations in the S protein of SARS-CoV-2; whereas, CoVariants only includes 12 mutations, and CoV-RDB focuses on neutralizing susceptibility to monoclonal antibodies, convalescent plasma, and vaccine plasma [24]. Besides, a database of human viruses transmitted via aerosols named AVM includes the immune escape of SARS-CoV-2 [25]. In 2023, Song *et al.* developed a comprehensive database named RCoV19, which included 12 554 entries about the mutations in the SARS-CoV-2 genome, assigned to six major aspects such as "infectivity/transmissibility," "antibody resistant," and "drug resistant" [26]. Despite all these databases, more detailed functional annotations about the mutations in the S protein of SARS-CoV-2 are still needed.

Here we presented an integrated and convenient query for annotations of mutations in the S protein of SARS-CoV-2. Based on 717 selected publications, we manually annotated the functions and/or effects of 205 single mutations and 93 multiple mutations in the S protein. We organized annotation entities using multilevel hierarchical categories and developed AnnCovDB, an annotation database for mutations in the S protein, with a searching and viewing interface.

Materials and methods

Data collection

Figure 1a shows a scheme for selecting amino acid mutations in the S protein of SARS-CoV-2 and searching for relevant publications. On 25 January 2024, we downloaded the dataset of amino acid mutations in the S protein involving 16 419 647 SARS-CoV-2 variants from the GISAID database (https://gisaid.org), using the wild-type sequence (Wuhan-hu-1 strain, GISAID accession EPI_ISL_402125) as reference. We selected mutations that are harbored in the S protein of at least 8000 SARS-CoV-2 variants. Then, on 8 November 2024, relevant publications in the PubMed database were searched using the keywords "SARS-CoV-2" AND "mutation [Title/Abstract]." The papers were filtered for journals with a ≥ 5 impact factor, obtaining 1071 publications. These publications were reviewed for descriptions of the functions or effects of S protein mutations, and annotation entities were manually curated based on 717 of them. If the conclusions of two publications on one aspect of a mutation differ, we reserve both. The data collection is performed every 3 months to include new publications for mutation annotation.

Multiple-level hierarchical organization of functional annotation entities

A large proportion of mutations contain multiple entities for the functional annotations. To categorize them, we introduce multilevel hierarchical categories, which are primarily adopted from the textbook "Principles of Virology" [27, 28]. The categories are divided into four levels of categories, as shown in Fig. 1b. The top-level categories are "Structure," "Infectious cycle," "Host–virus interaction," and "Pathogenesis and infection in animals." The "Structure" category includes the impact of mutations on the three-dimensional structure of the S protein as well as the stability. The "Infectious cycle" category records how mutations in the S protein affect SARS-CoV-2's attachment, entry, post-transcriptional, and assembly procession. The "Host–virus interaction" is divided into sub-categories "infectivity" and "immunity," including entities such as the effect of mutations on the immune escape of monoclonal antibodies and convalescent plasma. The "Pathogenesis" in "Pathogenesis and infection in animals" records the clinical manifestations of SARS-CoV-2 infection in humans. Considering the important role of animals in COVID-19 as natural survivors of SARS-CoV-2 [29, 30], we include "Infection in animal" in the 4th top category to record the viral mutations that occurred in animals and/or led to animal infections.

Web interface implementation

The front-end interface of the AnnCovDB was developed with JavaScript, a progressive JavaScript framework, and styled using Element UI (https://element.eleme.cn/) and Vuetify (https://vuetifyjs.com/en/), which provide a rich collection of user-interface components. Data visualization is implemented with ECharts (https://echarts.apache.org/en/). On the backend, Node.js (https://nodejs.org/en), a JavaScript runtime built on Chrome's V8 JavaScript engine, powers the server-side logic, while MongoDB serves (https://www.mongodb.com/) as the database for storing application data. For deployment, NGINX (https://nginx.org/en/) is used to serve the application to users efficiently. The AnnCovDB is freely available at https://AnnCovDB.app.bio-it.tech/.

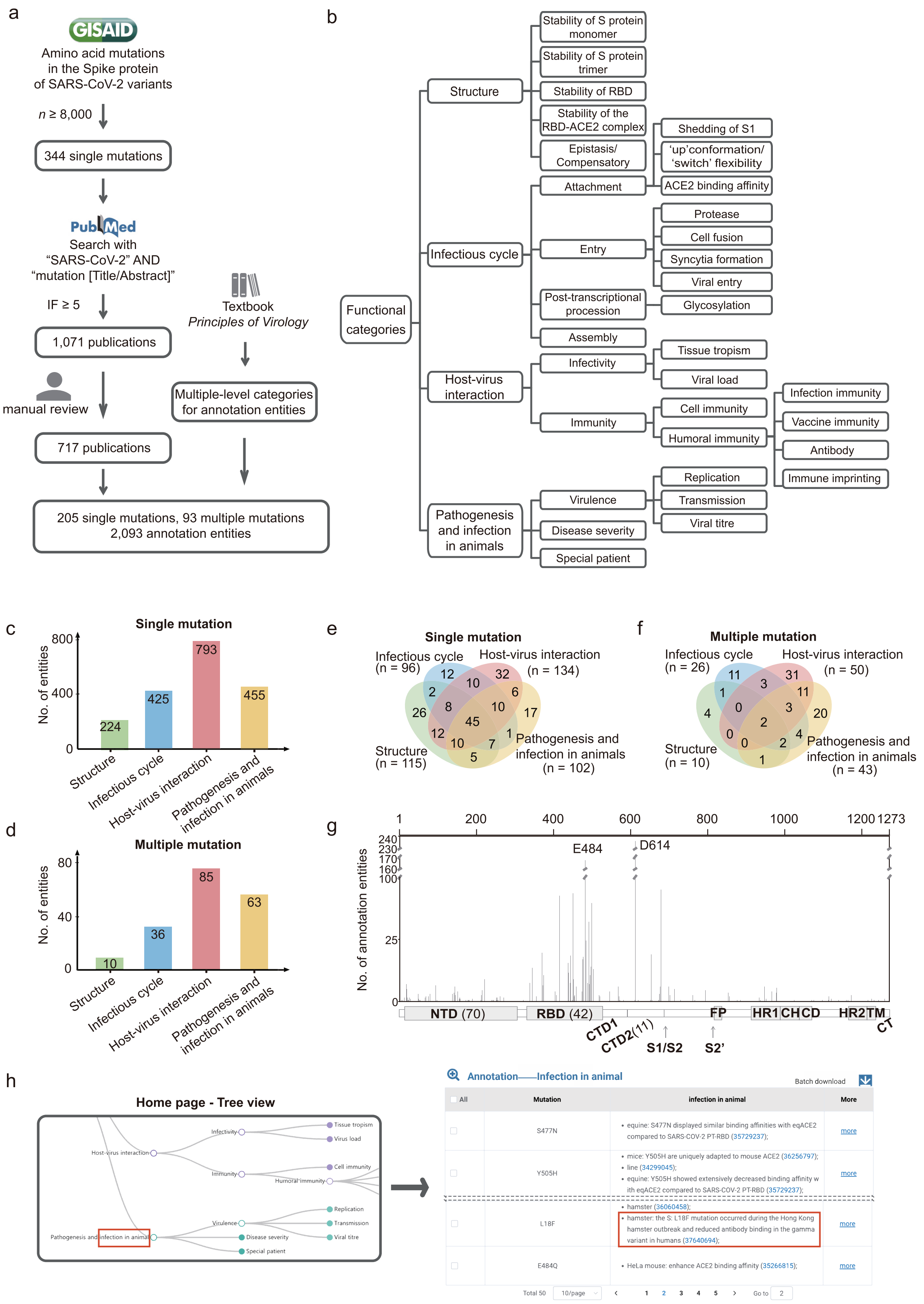
Results

Data statistics

Based on 717 publications, functional annotations for 205 single mutations and 93 multiple mutations in SARS-CoV-2 S protein were manually curated, generating a total of 2093 annotation entities. These entities were organized hierarchically into four-level categories (Fig. 1b). As shown in Fig. 1c and d, the top-level category "Host–virus interaction" has the most annotation entities, followed by "Pathogenesis and infection in animals," "Infectious cycle," and "Structure." About half of the single or multiple mutations (146 of 298, 49.0%) had more than one annotation category, and many of them were dispersed across multiple top-level categories, referred to as polyfunctional mutations. Of the 205 single mutations, 119 (58.0%) were polyfunctional (Fig. 1e); moreover, 45 (22.0%) contained annotation entities from four top-level categories. Notably, 31 of the 34 (91%) characteristic mutations identified in the Omicron BA.1 variants were polyfunctional [31].

For multiple mutations, 29.0% (27 of 93) are polyfunctional (Fig. 1f), with only two falling into four top-level categories (L452R/T478K and L452R/E484Q). Five mutations are prevalent in the combinations of multiple mutations, including D614G (within 33 multiple mutations), N501Y (25), E484K (17), S477N (11), and Q498R (10).

Sixty-seven (32.7%) single mutations are located in the receptor-binding domain (RBD, residues from 331 to 528) of the S protein, and 70 (34.1%) are in the N-terminal domain (NTD, residues from 14 to 306). RBD has 42 (45.2%) multiple mutations, while NTD has only one. As shown in Fig. 1g,



□ All	Mutation	infection in animal	More
□	S477N	• equine: S477N displayed similar binding affinities with eqACE2 compared to SARS-COV-2 PT-RBD (35729237);	more
□	Y505H	• mice: Y505H are uniquely adapted to mouse ACE2 (36256797); • line (34299045); • equine: Y505H showed extensively decreased binding affinity w ith eqACE2 compared to SARS-COV-2 PT-RBD (35729237);	more
□	L18F	• hamster (36060458); • hamster: the S: L18F mutation occurred during the Hong Kong hamster outbreak and reduced antibody binding in the gamma variant in humans (37640694);	more
□	E484Q	• HeLa mouse: enhance ACE2 binding affinity (35266815);	more

Figure 1. Overview of AnnCovDB. (a) A scheme for selecting mutations and publications for manual annotation curation. (b) Multiple-level hierarchical categories for organizing annotated entities of mutations. (c–d) The numbers of annotation entities of the four top-level categories for single mutations (c) and multiple mutations (d). (e–f) Venn diagrams of annotation entities divided into the four top-level categories for the single mutations (e) and multiple mutations (f). (g) The number of annotation entities in mutations located in the spike protein. The number of mutations in specific domains is listed in the brackets. (h) Screenshots of AnnCovDB pages for the browsing mutations associated with infection in animal.

the mutations in RBD and the S1/S2 cleavage site had remarkably more annotation entities. D614 and E484 are the two most extensively investigated mutations, with more than 100 annotation entities.

Applications

Searching functional annotations of "P681R" in the furin cleavage site

The characteristic mutation P681R in the S protein of B.1.617 (Delta) variants is well known for significantly increasing S1/S2 cleavage efficiency, resulting in faster viral replication [20]. However, it is less known that P681R is a polyfunctional mutation. Using the AnnCovDB database, users could obtain 45 annotation entities of P681R, which are spread in four top-level categories including "Infectious cycle" ($n = 20$), "Pathogenesis and infection in animals" ($n = 18$), "Host–virus interaction" ($n = 6$), and "Structure" ($n = 1$). For example, in the "Structure" category, P681R was predicted to decrease the stability of the S protein using the deep-learning-based method DeepDDG [32, 33]. In the "Host–virus interaction" category, the variants with P681R/D614G had a lower T-cell immunogenicity than those with simply D614G [34]. The combination of L452R/P681R/D950N had an annotation in "Pathogenesis and infection in animals" that was identified as essential for the higher ACE2 downregulation activity observed in the Delta variant compared to that in the other variants of concern [35].

Searching mutations and annotations related to SARS-CoV-2 infection in hamsters

Users could choose the "infection in animals" in the "Pathogenesis and infection in animals" top-level category from the interactive hierarchical categories of entities, and AnnCovDB would yield 40 single mutations and 12 multiple mutations related to this category (Fig. 1h). The annotation entity descriptions included animal species and identified nine mutations associated with SARS-CoV-2 infection in hamsters. H655Y, for example, has been associated with increased efficient transmission in a hamster infection model, possibly through enhanced S cleavage and viral growth [36]. On 14 October 2021, the L18F mutation emerged during the Hong Kong hamster outbreak [37]. Furthermore, neuroinvasion was associated with neuroinflammation in the olfactory bulb of hamsters inoculated with D614G [38].

Conclusion and discussion

Overall, AnnCovDB is a database of manually curated annotations for mutations in the S protein of SARS-CoV-2 with a high frequency. A total of 2093 annotation entities were organized into multiple-level hierarchical categories for users' convenience. This database allows researchers to more efficiently search for the functional effects or related underlying mechanisms of SARS-CoV-2 mutations. Additionally, considering the diminishing influence of COVID-19 on people and the subsequent decrease in the volume of new articles, this study plans to refresh the database every 3 months.

In some cases, the functional annotations about one mutation are inconsistent. We found that part could be attributed to the different methodologies. For instance, the mutation S371 F was described as enhancing the ACE2 affinity using the computational docking method [39], while in one deep mutational scanning dataset, it appears to decrease hACE2 affinity [40]. These inconsistencies were all recorded in AnnCovDB.

In the current version, the publications were queried in PubMed databases and filtered with certain criteria. We also tested using an artificial intelligence tool, ChatGPT, for the literature search. However, ChatGPT failed to provide a complete list of relevant publications. For example, for the Y505H mutation, ChatGPT v3.5 gave three publications and provided three annotations including "loss of hydrogen bonding," "role in immune evasion and infectivity," and "stability and structural Changes." In contrast, AnnCovDB contained 26 annotation entities for Y505H based on 24 publications. Recently, Lehr *et al.* reported that ChatGPT is a poor curator of scientific articles [41]. Therefore, manual curation might be labor-intensive but still useful approach. Thus, AnnCovDB uses the retrieval of the PubMed database and manually curated annotation.

The major limitation of AnnCovDB is that the annotations were largely dependent on the curators since AnnCovDB contained 29 categories for annotations. For a comparison, the databases CoV-RDB and RCoV19 have less than six annotation categories and the annotations are more standardized. It is a trade-off between the abundance of categories and standardization. More categories might provide more convenience for users to find the annotations of interest.

Conflict of interest: None declared.

Funding

This work was financially supported by Academy of Military Medical Sciences for innovation teams.

Data availability

All data in AnnCovDB are freely accessible at https://AnnCovDB.app.bio-it.tech/.

References

1. World Health Organization. "WHO COVID-19 Dashboard." WHO Data. Accessed June 6, 2024. https://data.who.int/dashboards/covid19/cases?n=c.
2. Gorbalenya AE, Baker SC, and Baric RS, Coronaviridae Study Group of the International Committee on Taxonomy of Viruses. The species Severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2. *Nat Microbiol* 2020;**5**:536–44. https://doi.org/10.1038/s41564-020-0695-z
3. Callaway E. Beyond Omicron: what's next for COVID's viral evolution. *Nature* 2021;**600**:204–07. https://doi.org/10.1038/d41586-021-03619-8
4. 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. https://doi.org/10.1038/s41580-021-00418-x
5. Hoffmann M, Kleine-Weber H, Schroeder S *et al.* SARS-CoV-2 cell entry depends on ACE2 and TMPRSS2 and is blocked by a clinically proven protease inhibitor. *Cell* 2020;**181**:271–280e8. https://doi.org/10.1016/j.cell.2020.02.052
6. 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. https://doi.org/10.1038/s41586-022-04474-x

7. Willett BJ *et al*. SARS-CoV-2 Omicron is an immune escape variant with an altered cell entry pathway. *Nat Microbiol* 2022;**7**:1161–79.
8. Yao Z, Zhang L, Duan Y *et al*. Molecular insights into the adaptive evolution of SARS-CoV-2 spike protein. *J Infect* 2024;**88**:106121. https://doi.org/10.1016/j.jinf.2024.106121
9. Isabel S, Graña-Miraglia L, Gutierrez JM *et al*. Evolutionary and structural analyses of SARS-CoV-2 D614G spike protein mutation now documented worldwide. *Sci Rep* 2020;**10**:14031. https://doi.org/10.1038/s41598-020-70827-z
10. Wrapp D, Wang N, Corbett KS *et al*. Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. *Science* 2020;**367**:1260–63. https://doi.org/10.1126/science.abb2507
11. Zhang L, Jackson CB, Mou H *et al*. SARS-CoV-2 spike-protein D614G mutation increases virion spike density and infectivity. *Nat Commun* 2020;**11**:6013. https://doi.org/10.1038/s41467-020-19808-4
12. Plante JA, Liu Y, Liu J *et al*. Spike mutation D614G alters SARS-CoV-2 fitness. *Nature* 2021;**592**:116–21. https://doi.org/10.1038/s41586-020-2895-3
13. Zhang J, Cai Y, Xiao T *et al*. Structural impact on SARS-CoV-2 spike protein by D614G substitution. *Science* 2021;**372**:525–30. https://doi.org/10.1126/science.abf2303
14. Kumar S, Thambiraja TS, Karuppanan K *et al*. Omicron and Delta variant of SARS-CoV-2: a comparative computational study of spike protein. *J Med Virol* 2022;**94**:1641–49. https://doi.org/10.1002/jmv.27526
15. Kumar S, Karuppanan K, Subramaniam G. Omicron (BA.1) and sub-variants (BA.1.1, BA.2, and BA.3) of SARS-CoV-2 spike infectivity and pathogenicity: a comparative sequence and structural-based computational assessment. *J Med Virol* 2022;**94**:4780–91. https://doi.org/10.1002/jmv.27927
16. Liu Y, Liu J, Plante KS *et al*. The N501Y spike substitution enhances SARS-CoV-2 infection and transmission. *Nature* 2022;**602**:294–99. https://doi.org/10.1038/s41586-021-04245-0
17. Moulana A, Dupic T, Phillips AM *et al*. Compensatory epistasis maintains ACE2 affinity in SARS-CoV-2 Omicron BA.1. *Nat Commun* 2022;**13**:7011.https://doi.org/10.1038/s41467-022-34506-z
18. Callaway E. The mutation that helps Delta spread like wildfire. *Nature* 2021;**596**:472–73. https://doi.org/10.1038/d41586-021-02275-2
19. Mlcochova P, Kemp SA, Dhar MS *et al*. SARS-CoV-2 B.1.617.2 Delta variant replication and immune evasion. *Nature* 2021;**599**:114–19. https://doi.org/10.1038/s41586-021-03944-y
20. Saito A, Irie T, Suzuki R *et al*. Enhanced fusogenicity and pathogenicity of SARS-CoV-2 Delta P681R mutation. *Nature* 2022;**602**:300–06. https://doi.org/10.1038/s41586-021-04266-9
21. Flemming A. Omicron, the great escape artist. *Nat Rev Immunol* 2022;**22**:75. https://doi.org/10.1038/s41577-022-00676-6
22. Callaway E. Omicron likely to weaken COVID vaccine protection. *Nature* 2021;**600**:367–68. https://doi.org/10.1038/d41586-021-03672-3
23. Feng Y, Yi J, Yang L *et al*. COV2Var, a function annotation database of SARS-CoV-2 genetic variation. *Nucleic Acids Res* 2024;**52**:D701–D713. https://doi.org/10.1093/nar/gkad958
24. Tzou PL, Tao K, Pond SLK *et al*. Coronavirus Resistance Database (CoV-RDB): SARS-CoV-2 susceptibility to monoclonal antibodies, convalescent plasma, and plasma from vaccinated persons. *PLoS One* 2022;**17**:e0261045. https://doi.org/10.1371/journal.pone.0261045
25. Mei L, Hou Y, Zhou J *et al*. AVM: a manually curated database of aerosol-transmitted virus mutations, human diseases, and drugs. *Genom Proteom Bioinform* 2024;**22**:qzae041. https://doi.org/10.1093/gpbjnl/qzae041
26. Li C, Ma L, Zou D *et al*. RCoV19: a one-stop hub for SARS-CoV-2 genome data integration, variant monitoring, and risk prewarning. *Genom Proteom Bioinform* 2023;**21**:1066–79. https://doi.org/10.1016/j.gpb.2023.10.004
27. Flint SJ, Enquist LW, Racaniello VR, Skalka AM. *Principles of Virology*, volume I: Molecular Biology. 3rd edn. ASM Press, 2015.
28. Flint SJ, Enquist LW, Racaniello VR, Skalka AM. *Principles of Virology*, volume II: Pathogenesis and Control. 3rd edn. ASM Press, 2015.
29. Shi J, Wen Z, Zhong G *et al*. Susceptibility of ferrets, cats, dogs, and other domesticated animals to SARS-coronavirus 2. *Science* 2020;**368**:1016–20. https://doi.org/10.1126/science.abb7015
30. Mahdy MAA, Younis W, Ewaida Z. An overview of SARS-CoV-2 and animal infection. *Front Vet Sci* 2020;**7**:596391. https://doi.org/10.3389/fvets.2020.596391
31. Viana R, Moyo S, Amoako DG *et al*. Rapid epidemic expansion of the SARS-CoV-2 Omicron variant in southern Africa. *Nature* 2022;**603**:679–86. https://doi.org/10.1038/s41586-022-04411-y
32. Cao H, Wang J, He L *et al*. DeepDDG: predicting the stability change of protein point mutations using neural networks. *J Chem Inf Model* 2019;**59**:1508–14. https://doi.org/10.1021/acs.jcim.8b00697
33. Zeng L, Lu Y, Yan W *et al*. A protein co-conservation network model characterizes mutation effects on SARS-CoV-2 spike protein. *Int J Mol Sci* 2023;**24**:3255. https://doi.org/10.3390/ijms24043255
34. Boon SS, Xia C, Wang MH *et al*. Temporal-geographical dispersion of SARS-CoV-2 spike glycoprotein variant lineages and their functional prediction using in silico approach. *mBio* 2021;**12**:e0268721. https://doi.org/10.1128/mBio.02687-21
35. Maeda Y, Toyoda M, Kuwata T *et al*. Differential ability of spike protein of SARS-CoV-2 variants to downregulate ACE2. *Int J Mol Sci* 2024;**25**:1353. https://doi.org/10.3390/ijms25021353
36. Rathnasinghe R, Jangra S, Ye C *et al*. Characterization of SARS-CoV-2 Spike mutations important for infection of mice and escape from human immune sera. *Nat Commun* 2022;**13**:3921. https://doi.org/10.1038/s41467-022-30763-0
37. McBride DS, Garushyants SK, Franks J *et al*. Accelerated evolution of SARS-CoV-2 in free-ranging white-tailed deer. *Nat Commun* 2023;**14**:5105. https://doi.org/10.1038/s41467-023-40706-y
38. Bauer L, Rissmann M, Benavides FFW *et al*. In vitro and in vivo differences in neurovirulence between D614G, Delta And Omicron BA.1 SARS-CoV-2 variants. *Acta Neuropathol Commun* 2022;**10**:124. https://doi.org/10.1186/s40478-022-01426-4
39. Singh P, Sharma K, Shaw D *et al*. Mutational characterization of Omicron SARS-CoV-2 lineages circulating in Chhattisgarh, a central state of India. *Front Med Lausanne* 2023;**9**:1082846. https://doi.org/10.3389/fmed.2022.1082846
40. Javanmardi K, Segall-Shapiro TH, Chou C-W *et al*. Antibody escape and cryptic cross-domain stabilization in the SARS-CoV-2 Omicron spike protein. *Cell Host Microbe* 2022;**30**:1242–1254.e6. https://doi.org/10.1016/j.chom.2022.07.016
41. Lehr SA, Caliskan A, Liyanage S *et al*. ChatGPT as research scientist: probing GPT's capabilities as a research librarian, research ethicist, data generator, and data predictor. *Proc Natl Acad Sci U S A* 2024;**121**:e2404328121. https://doi.org/10.1073/pnas.2404328121

Database, 2025, 00, baaf002, DOI: https://doi.org/10.1093/database/baaf002, Database tool

© The Author(s) 2025. Published by Oxford University Press.
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.